EFSA BIOHAZ Panel (EFSA Panel on Biological Hazards) , 2013 . Scientific Opinion on the evaluation of molecular typing methods for major food-borne microbiological hazards and their use for attribution modelling, outbreak investigation and scanning surveillance: Part 1 (evaluation of methods and applications)

EFSA Publication

Link to article, DOI:
10.2903/j.efsa.2013.3502

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

SCIENTIFIC OPINION

Scientific Opinion on the evaluation of molecular typing methods for major food-borne microbiological hazards and their use for attribution modelling, outbreak investigation and scanning surveillance: Part 1 (evaluation of methods and applications)[1]

EFSA Panel on EFSA Biological Hazards (BIOHAZ)[2, 3]

European Food Safety Authority (EFSA), Parma, Italy

ABSTRACT

An evaluation of molecular typing methods that can be applied to the food-borne pathogens *Salmonella*, *Campylobacter*, Shiga toxin-producing *Escherichia coli* and *Listeria monocytogenes* is presented. This evaluation is divided in two parts. Firstly, commonly used molecular typing methods are assessed against a set of predefined criteria relating to discriminatory capacity, reproducibility, repeatability and current or potential suitability for international harmonisation. Secondly, the methods are evaluated for their appropriateness for use in different public health-related applications. These applications include outbreak detection and investigation, attribution modelling, the potential for early identification of food-borne strains with epidemic potential and the integration of the resulting data in risk assessment. The results of these evaluations provide updated insights into the use and potential for use of molecular characterisation methods, including whole genome sequencing technologies, in microbial food safety. Recommendations are also made in order to encourage a holistic and structured approach to the use of molecular characterisation methods for food-borne pathogens; in particular, on the importance of structured co-ordination at international level to help overcome current limitations in harmonisation of data analysis and interpretation.

KEY WORDS

genotyping, molecular typing, whole genome sequencing, outbreak, source attribution, epidemic potential

---

[1] On request from EFSA, Question No EFSA-Q-2013-00032, adopted on 5 December 2013.

[2] Panel members: Olivier Andreoletti, Dorte Lau Baggesen, Declan Bolton, Patrick Butaye, Paul Cook, Robert Davies, Pablo S. Fernandez Escamez, John Griffin, Tine Hald, Arie Havelaar, Kostas Koutsoumanis, Roland Lindqvist, James McLauchlin, Truls Nesbakken, Miguel Prieto Maradona, Antonia Ricci, Giuseppe Ru, Moez Sanaa, Marion Simmons, John Sofos and John Threlfall. Correspondence: biohaz@efsa.europa.eu

[3] Acknowledgement: The Panel wishes to thank the members of the Working Group on the evaluation of molecular typing methods for major food-borne pathogens: Dorte Lau Baggesen, Patrick Butaye, Robert Davies, Tine Hald, Arie Havelaar, Bjørn-Arne Lindstedt, Martin Maiden, Eva Møller Nielsen, Gaia Scavia and John Threlfall for the preparatory work on this scientific opinion and European Centre for Disease Prevention and Control (ECDC) staff: Marc Struelens, and EFSA staff: Luis Vivas-Alegre, Ernesto Liebana Criado and Maria Teresa da Silva Felicio for the support provided to this scientific opinion.

Suggested citation: EFSA BIOHAZ Panel (EFSA Panel on Biological Hazards), 2013. Scientific Opinion on the evaluation of molecular typing methods for major food-borne microbiological hazards and their use for attribution modelling, outbreak investigation and scanning surveillance: Part 1 (evaluation of methods and applications). EFSA Journal 2013;11(12):3502, 84 pp. doi:10.2903/j.efsa.2013.3502

Available online: www.efsa.europa.eu/efsajournal

SUMMARY

The European Food Safety Authority (EFSA) asked the Panel on Biological Hazards (BIOHAZ) to deliver a scientific opinion on the evaluation of molecular typing methods for major food-borne microbiological hazards and their use for attribution modelling, outbreak investigation and scanning surveillance. In particular, this opinion addresses the first two terms of reference of the mandate, namely: (i) to review information on current and prospective (e.g. whole genome sequencing (WGS)) molecular characterisation and sub-typing methods for food-borne pathogens (e.g. *Salmonella*, *Campylobacter*, Shiga toxin-producing *Escherichia coli* (STEC) and *Listeria*) in terms of discriminatory capability, reproducibility, and capability for international harmonisation, and (ii) to review the appropriateness of use of the different food-borne pathogen sub-typing methodologies (including data analysis methods) for outbreak investigation, attribution modelling and the potential for early identification of food-borne organisms with epidemic potential and the integration of the resulting data in risk assessment.

In the approach taken by the BIOHAZ Panel to the reply to these two terms of reference, the starting point is a bacterial isolate from a human, food, animal or environmental source which has already been characterised to genus or species level. The BIOHAZ Panel acknowledged that in the future, bacterial identification and molecular typing may be combined in a single procedure and included in a culture-independent diagnostic process. There is very little relevant experience regarding the application of such metagenomic approaches in the food-borne zoonoses field and therefore this area is not considered in this Opinion.

The BIOHAZ Panel highlights that all bacteria are subject to genetic change (e.g. in response to environmental stress and human interventions such as antimicrobial or heavy metal use or vaccination), sometimes by mutation but more often by acquisition or loss of genetic elements. These changes can be followed by clonal expansion in the case of biologically successful organisms. Ongoing evolution driven by genetic change and selection has given rise to highly adaptable organisms that are able to exploit and expand into novel niches and extend their host range. Such evolution may also be linked to the emergence of various 'epidemic' strains of pathogens, such as *Salmonella*, in combination with other biological factors and epidemiological opportunities for dissemination. The molecular characteristics of organisms provide markers for investigation of outbreaks, attribution studies, and assessment of potential virulence or epidemic potential. The BIOHAZ Panel also points out that even with high-resolution molecular approaches, up to and including WGS analysis, it is not possible to establish how closely two isolates are related without an appreciation of the structure and diversity of the bacterial population in question. Further, to properly evaluate typing methodologies, data from strain characterisation should be linked with epidemiological metadata and the strain selection must be unbiased and statistically representative of the population to be assessed. International harmonisation of molecular characterisation outputs by means of standardisation or appropriate quality control procedures is essential. This includes controlling the accuracy of production of DNA sequences from WGS and the further interpretations of annotation pipelines.

For the evaluation of molecular typing methods, the BIOHAZ Panel established a set of pre-defined criteria based on the first term of reference. These criteria included: (i) discriminatory capacity (i.e. degree of discrimination between strains of different genotype), (ii) reproducibility and repeatability (i.e. consistency of results within and between laboratories, and over time), (iii) current international harmonisation (i.e. status with regard to availability and use of standard operational procedures, external quality assurance systems, harmonised nomenclature and data management tools), and (iv) the potential for future international harmonisation in situations where any of the sub-criteria under (iii) may not be currently harmonised.

Following the evaluation against those criteria, the BIOHAZ Panel concluded that molecular typing methods should ideally provide appropriate discriminatory power, reproducibility, capability for international harmonisation and reduced handling of and exposure to pathogens in the laboratories. No

current typing method, whether phenotypic or molecular, complies with all these expectations. Several methods are often used in combination in order to obtain the resolution needed. The methods applied depend on the pathogen and on the application sought. These methods have proven track records of use, and for some of them (e.g. Multi locus sequence typing (MLST), Pulsed-field gel electrophoresis (PFGE)) extensive databases of valuable typing data have been collected. Further, methods based on WGS can replace and are increasingly replacing the numerous different methodologies currently in use in human and veterinary reference laboratories, and the same methods can be used for all organisms. An essential precondition is the availability of quality control methods, to ensure the reliability and consistency of molecular data generated, coupled with high quality bioinformatics support for the analysis of the data generated. The BIOHAZ Panel acknowledged that, regarding WGS, limited knowledge is available in relation to the technical errors that occur during sequencing and analysis and on the effect of genetic drift in the different bacterial populations over time, which may complicate the interpretation of results.

With regard to the review of the appropriateness of use of the different food-borne pathogen sub-typing methodologies for different food-safety related public health applications (i.e. detection and investigation of food-borne outbreaks of disease, food-borne source-attribution, early identification of food-borne organism with epidemic potential and their integration in risk assessment) the BIOHAZ Panel concluded that detection of outbreaks and their investigation in real-time would be enhanced by the generation of fully comparable molecular typing data from human, veterinary and food laboratories prior to submission to a central or connected databases. Some molecular typing methods (e.g. MLST, PFGE, Multi locus variable tandem repeat analysis (MLVA)) have been harmonised to a greater or lesser extent for the purpose of outbreak detection and investigation. The international development of harmonised platforms for WGS-generated data should be encouraged.

In relation to source-attribution analysis of food-borne pathogens, the Panel concluded that a major challenge of using data generated from molecular typing methods in source attribution models, in particular WGS data, will be to define meaningful subtypes providing an appropriate level of discrimination for source attribution. A high level of discrimination is not necessarily the best option. The applied method has to allow for some genetic diversity between isolates from human and animal/food sources, but only to the degree so that it can still be assumed that they originate from the same source. Independent of the choice of molecular typing method and approach for source attribution, it is important that the data included from human and potential sources are related in time and space. Source attribution analysis is, therefore, facilitated by integrated surveillance providing a collection of isolates from all (major) sources that should, to the extent possible, represent what the human population is exposed to.

In relation to the last of the applications, the BIOHAZ Panel concluded that the epidemic potential of a food-borne strain within a bacterial species, or even within a subtype varies considerably, and is a function of its inherent genetic characteristics and their expression combined with ecological factors including the opportunities to spread in the food chain. Prediction of the public health risk and epidemic potential of emerging strains of food-borne pathogens has not yet been possible. Nevertheless, if an epidemic strain has already emerged in a certain region such a strain can be rapidly characterised employing current molecular typing methods and thus serve to identify the occurrence of such strains in other regions for risk management purposes. High throughput WGS technologies offer new opportunities to characterise bacterial strains in great detail. The genetic information that these technologies provide will however need to be considered together with gene expression, host and ecological factors, including the opportunities to spread in the food chain. Finally, although there are differences between bacterial species, the principle of assessing the gene content in relation to fitness as a means to assess risk potential that has been used for the four organisms considered in this opinion should be applicable to any bacteria.

The BIOHAZ Panel makes a series of recommendations on important issues to be considered as these methods, in particular WGS analysis, have limitations when using the data they generate. Thus, modern molecular typing methods provide many opportunities for rapid and accurate determination of

the genealogical relationships among bacterial isolates. Interpretation of the results generated by these methods for different public health applications requires this information to be placed in the context of the diversity, degree of genetic change (e.g. during storage of isolates or mutation during an outbreak and in reservoirs) and population structure of the particular pathogen in question. Therefore, large scale carefully co-ordinated studies are required to fully elucidate this. The development of more informative and easier to use bioinformatic tools for analysis of WGS data is needed. Multidisciplinary and integrated research programs are needed to develop and validate the use of detailed genetic information for 'predictive' hazard identification, accounting for gene expression and how this affects the fate of pathogens in the food chain and their interaction with human and animal hosts. Further recommendations are made on particular issues to aid the use of these methods and the data they generate for the different applications considered.

TABLE OF CONTENTS

Abstract 1
Summary 2
Table of contents 5
Background as provided by EFSA 7
Terms of reference as provided by EFSA 8
Assessment 9
1. Introduction 9
2. Bacterial populations: structure and epidemiology 12
2.1. Introduction 12
2.1.1. Relationships of pathogen population structure and epidemiology to molecular epidemiology 12
2.1.2. Forces that structure bacterial populations 12
2.2. *Campylobacter* spp. 15
2.3. *Salmonella* spp. 16
2.4. *Escherichia coli* 18
2.4.1. The concepts of STEC, EHEC and EAEC 18
2.4.2. Virulence attributes of STEC and EAEC 18
2.4.3. The 2011 *E. coli* O104:H4 outbreak 19
2.5. *Listeria monocytogenes* 19
2.6. Concluding remarks on relationships of pathogen population structure and epidemiology to molecular epidemiology 20
3. Review of current and prospective molecular identification and typing methods 23
3.1. Current methods 24
3.1.1. Molecular serotyping 24
3.1.2. Restriction Fragment Length Polymorphism (RFLP) analysis 25
3.1.3. Pulsed-Field Gel Electrophoresis (PFGE) analysis 25
3.1.4. Specific gene characterisation 26
3.1.5. Multiple-Locus Variable number tandem repeat Analysis (MLVA) 27
3.1.6. Sequence-based typing methods 28
3.2. Novel and prospective technologies 29
3.2.1. Whole genome mapping (WGM) analysis 29
3.2.2. Whole Genome Sequence (WGS) analysis 29
3.3. Concluding remarks on the review of the molecular typing methods 31
4. The use of molecular typing methods for the detection and investigation of food-borne outbreaks of disease 34
4.1. Introduction 34
4.2. Detection of food-borne outbreaks 34
4.3. Investigation of food-borne outbreaks 35
4.4. Data needs 35
4.5. Past and current experiences employing molecular typing methods 36
4.6. Current molecular methods used for outbreak detection and outbreak investigation 37
4.6.1. *Salmonella* spp. 37
4.6.2. *Listeria* 37
4.6.3. STEC 38
4.6.4. *Campylobacter* 38
4.6.5. Antibiotic resistance genes, virulence, and associated elements 39
4.7. Potential use of new technologies 39
4.8. Concluding remarks 39
5. The use of molecular typing methods for food-borne source attribution 40
5.1. Requirements and role of typing for food-borne hazards source attribution 40
5.1.1. The microbial subtyping approach 41
5.1.2. Available source attribution models 42
5.2. Optimal data needs 44

5.3. Current experiences in source attribution employing molecular sub-typing methods 45
5.4. Potential use of new technologies 46
5.5. Concluding remarks on the use of molecular typing methods in food-borne source attribution 47
6. The use of molecular typing methods in the early identification of food-borne organisms with epidemic potential and their integration in risk assessment 49
6.1. Data needs for the identification of food-borne organisms with epidemic potential 49
6.2. Future perspectives and potential use of new technologies 54
6.3. Concluding remarks on use of molecular typing methods in the early identification of food-borne organisms with epidemic potential and their integration in risk assessment 55
Conclusions and recommendations 57
References 60
Appendices 78
Appendix A. Summary of the evaluation of molecular typing methods for *Salmonella* spp., STEC, *Listeria monocytogenes* and *Campylobacter* spp. 78
Glossary 82
Abbreviations 84

BACKGROUND AS PROVIDED BY EFSA

It is important to link closely molecular surveillance initiatives instigated for pathogens identified in the human population and surveillance activities in food, feed and food-producing animals. This would help to identify common sources of infection for the animals themselves, e.g. via internationally-traded feed ingredients and replacement breeding and commercial stock, and would provide a means of comparing human and animal strains via real time surveillance and as part of outbreak investigations.

A wide variety of sub-typing methods exist for most pathogens but they are often applied in a way that is not standardised and dependent on individual protocols, approaches and equipment used in separate laboratories. The introduction of harmonised protocols and reference strains e.g. for pulsed field gel electrophoresis (PFGE), and for Multiple-Locus Variable Number Tandem Repeat Analysis (MLVA) as part of the PulseNet[4] initiative represent an attempt to introduce harmonisation of methodology or standardisation of interpretation. PulseNet in particular has been particularly valuable in the USA, identifying numerous diffuse common source outbreaks of *Salmonella* spp. or STEC[5] that would otherwise have been considered to be sporadic cases. The identification of such outbreaks allows interventions such as product recall that can shorten the duration of food-borne disease outbreaks and potentially save lives. Furthermore, by identifying the factors that caused the outbreak, HACCP plans and food safety standards may be reviewed, helping to reduce future outbreaks or sporadic cases.

In recent years EFSA has made increasing use of attribution modelling to enhance the scientific value of opinions. This approach has been very valuable to help risk managers focus regulatory attention on the highest priority sources of food-borne infection. The precision of attribution modelling based on sub-typing of organisms is limited both by the scarcity of harmonised data for some food animal species, e.g. for *Salmonella* spp. in the bovine reservoir and the occurrence of similar organisms at the serovar level in different animal populations. In the case of other organisms such as *Campylobacter*, even this level of sub-typing detail is largely lacking. Various studies have shown that in many cases further distinction between sources, both in terms of animal reservoir and geographical origin can be made by inclusion of additional combinations of phenotypic or molecular sub-typing data. A notable example of this is the use of multi-locus sequence typing (MLST) for *Campylobacter* in studies in New Zealand and UK. It has recently been demonstrated that the use of MLST typing data in combination with case-control studies provides novel perspectives on the risk factors for human disease in relation to different animal reservoirs.

For the future, sequence-based approaches, including whole genome sequencing (WGS), have prepared the stage for revolutionary advances in diagnostic and typing techniques. Increasing use of data generated from next-generation sequencing (NGS) technologies is expected to provide the means for a paradigm shift in the way microorganisms are identified and compared. This will result in greater ability to undertake detailed analysis and more rapidly identify dispersed outbreaks such as those arising from contaminated foods. Epigenetic techniques and quantitative gene expression arrays may also in the future be used to provide early indication of potential new and emerging epidemic strains.

Harmonised approaches for (i) selection of representative isolates of food-borne pathogens, (ii) selection of sub-typing methodologies, and (iii) analysis and storage of large quantities of molecular typing data, would provide valuable guidance from EFSA to the scientific community and regulatory bodies, specifically in the areas of outbreak detection and source attribution modelling for food-borne pathogens. To that end it is the intention to request participation of ECDC and EU Reference laboratories in this working group. Such an approach would enhance the value and integration of current molecular typing schemes and should ultimately assist in the application of improved tools to further enhance the protection of public health.

[4] Further information on PulseNet international available at: http://www.pulsenetinternational.org/ (last visited on 11/12/2013)

[5] Shiga toxin-producing *E. coli*

TERMS OF REFERENCE AS PROVIDED BY EFSA

EFSA requests the BIOHAZ Panel to:

1. Review information on current and prospective (e.g. WGS) molecular identification and sub-typing methods for food-borne pathogens (e.g. *Salmonella*, *Campylobacter*, STEC and *Listeria*) in terms of discriminatory capability, reproducibility, and capability for international harmonisation.

2. Review the appropriateness of use of the different food-borne pathogen sub-typing methodologies (including data analysis methods) for outbreak investigation, attribution modelling and the potential for early identification of organisms with future epidemic potential.

3. Evaluate the requirements for the design of surveillance activities for food-borne pathogens, in particular for the selection for a statistically representative group of isolates to be included in molecular typing investigations, and attribution modelling.

4. Review the requirements for harmonised data collection, management and analysis, with the final aim to achieve full integration of efficient and effectively managed molecular typing databases for food-borne pathogens.

Following a proposal made by the BIOHAZ Panel, EFSA agreed upon the delivery of two separate scientific Opinions: one covering Terms of reference one and two (deadline December 2013), and another covering Terms of reference three and four (deadline July 2014).

ASSESSMENT

1. Introduction

Molecular typing can be defined as the classification of microorganisms on the basis of variation in the genotype, and/or the presence or absence of specific genes (such as those which may contribute to the pathogenicity of the organism or to its ability to survive in less favourable environments) (Hallin et al., 2012). 'Genotype' has been defined as the genetic constitution of an organism, as assessed by a molecular method (van Belkum et al., 2007).

According to the European Centre for Disease Prevention and Control (ECDC), molecular typing refers to the application of laboratory methods capable of characterising, discriminating and indexing subtypes of microorganisms. Molecular typing of pathogens that cause infectious diseases complements traditional epidemiological surveillance by providing appropriate discriminatory analyses to allow the rapid and early detection of outbreaks, to detect and investigate transmission chains and the relatedness of strains, and to detect the emergence of antimicrobial resistance and new evolving pathogenic strains. Molecular typing can also support studies to trace-back the source of an outbreak and identify new risk factors, by linking isolates more accurately to epidemiological and clinical data (ECDC, 2007 and 2013).

The high degree of genetic structuring present in bacterial populations is the basis of molecular epidemiological studies of their distribution and spread. This population structure is reflected in hierarchical nomenclature schemes that group bacteria into orders, families, genera, species and subspecies. This taxonomic classification is governed by the International Code of Nomenclature of Bacteria (Lapage et al., 1992). Other nomenclature employed below the species level is less well defined and varies between different genera (van Belkum et al., 2007). As an example of this, Figure 1 below provides a general indication of the nomenclature employed to illustrate the transition between 'formal' taxonomic nomenclature and molecular typing-based nomenclature when employing multilocus sequence typing (MLST) (Maiden et al., 2013). The latter shows the different names usually employed depending on the level of discrimination between isolates achieved.

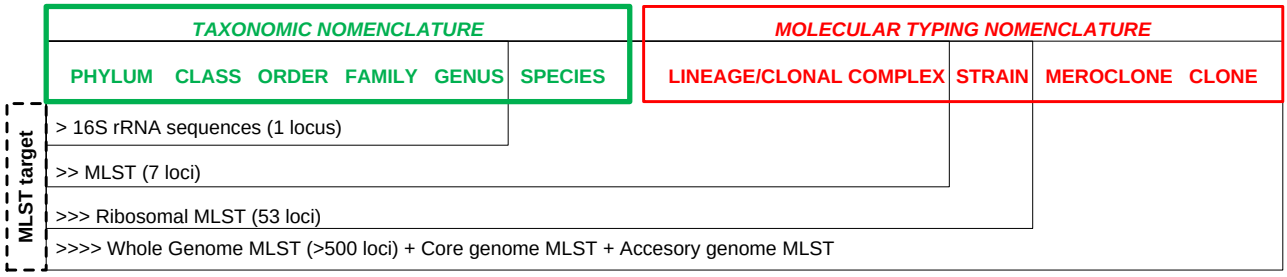

Figure 1: Example of taxonomic nomenclature and general molecular typing nomenclature based on level of discrimination between isolates achieved when employing multi locus sequence typing (MLST) (Modified from Maiden et al., 2013).

The significance of genetic structuring for public health is two-fold: (i) different subgroups of bacteria, even within species, can vary widely in their phenotypic properties, including those related to pathogenicity, such as virulence or host association; (ii) the size and diversity of bacterial populations is such that it is necessary to be able to distinguish variants within of isolates for the purposes of epidemiological analysis and, particularly, in the context of food-borne infections and source tracing.

Molecular epidemiology can be used to assess the contribution of genetic factors to the aetiology, epidemiology, and prevention of disease across populations. This model is analogous to that of traditional and clinical epidemiology - i.e. to investigate disease prevalence and incidence with respect to exposure to various risk factors, and to identify genes and genetic elements that contribute to

disease. The main practical purpose of molecular epidemiology is to generate or test hypotheses as to whether cases of infectious diseases are linked by recent transmission events or acquired from a potential source of contamination by comparing molecular typing data to relevant epidemiological data. Another key purpose is to explain how virulence and other phenotypic traits evolve in microorganisms over time, and thereby contribute to the survival of the organism and to disease severity and spread.

The scope of this Scientific Opinion is to evaluate current and prospective molecular methods for the epidemiological typing of specific food-borne pathogens (i.e. *Salmonella* spp., *Campylobacter* spp. (thermophilic), pathogenic Shiga toxin-producing *E. coli* (STEC[6]) and *Listeria monocytogenes*), and their potential application as tools in different food-safety related aspects as follows: (i) outbreak detection and investigation, (ii) source attribution, (iii) early identification of organisms with epidemic potential and risk assessment. An assessment of the different molecular methods based on three main criteria (i.e. discriminatory capability, reproducibility, and capability for international harmonisation) is presented with the intention to support decision making.

The starting point in the context of this Opinion is a bacterial isolate from a human, food, animal or environmental source which has already been characterised to genus or species level. In the future, bacterial identification and molecular typing may be combined in a single procedure and included in a culture-independent diagnostic process. There is very little relevant experience regarding the application of such metagenomic approaches in the food-borne zoonoses field and therefore this area is not considered in this Opinion.

Molecular identification and subtyping methods can be evaluated either from a hazard, or from an application perspective (Figure 2). Method-hazard combinations may have differing advantages and disadvantages depending on the final application sought.

[6] Also known as verotoxigenic *E. coli*, verocytotoxigenic *E. coli*, verotoxin-producing *E. coli* and verocytotoxin-producing *E. coli* (VTEC).

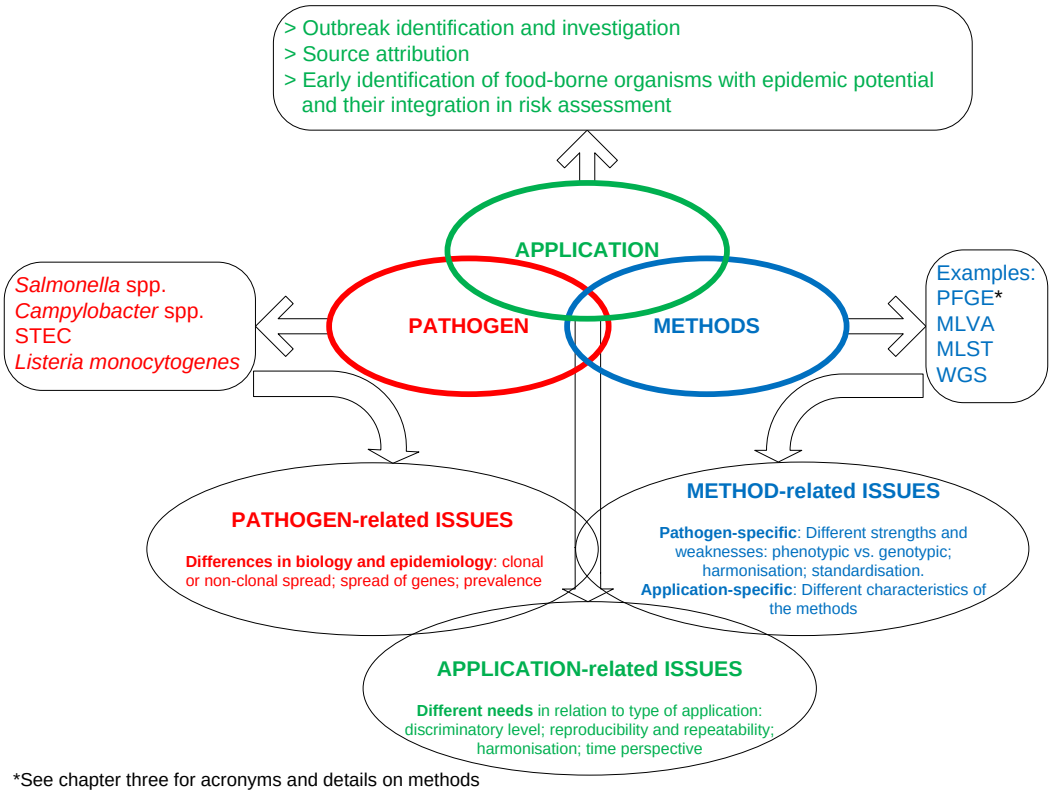

Figure 2: The multi-perspective approach considered when addressing the terms of reference of the mandate. General considerations on the pathogen-method-application perspectives and their possible correlations are described.

In the context of this Opinion, the application perspective was considered to provide the most useful evaluation for risk managers, as this is linked to the intended end–point use and not only with the characterisation of an organism. Secondary to the application perspective, the document is structured to take into account the method and/or the hazard addressed. The aim of this approach is to be able to present a comparative assessment that should aid the selection of the most suitable methods for different purposes, considering that these may vary depending on the hazard.

Recommendations are made taking into account practicalities when implementing different methods. In this context, wide access to equipment and availability of personnel with the appropriate skills are extremely important. Specific comments are made regarding the complexity of the implementation of the methodology, interpretation of the results obtained and on other practical considerations.

As indicated above, after the terms of reference proposed by EFSA, the requirements for the design of surveillance activities for food-borne pathogens and the principles and requirements for harmonised data collection, management and analysis will be evaluated and reviewed in a separate opinion to be delivered by mid 2014. This will be done in parallel to the ongoing EFSA and ECDC activities on the establishment of molecular typing databases[7].

[7] For further information on the mandate from the European Commission to EFSA and ECDC on the building of a molecular typing database visit: http://registerofquestions.efsa.europa.eu/roqFrontend/questionLoader?question=EFSA-Q-2013-00250 (last visited on 11/12/2013)

2. Bacterial populations: structure and epidemiology

2.1. Introduction

2.1.1. Relationships of pathogen population structure and epidemiology to molecular epidemiology

To establish the epidemiological relationship of any two isolates of a particular pathogen successfully it is necessary, not only to type them at a sufficient level of resolution, but also to place this in the context of the variation of the population as whole. This builds on previous knowledge of taxonomic relations between genera, species, subspecies, clades and subtypes (Philipott et al., 2010). Many typing methods have developed following the discovery of reagents that discriminate among pathogen variants (Struelens, 1998). To aid this effort a series of nomenclature schemes have been developed, often based on the technique used to discriminate isolates; hence 'serotypes' refer to a group of isolates which all react with the same immunological reagent. These were originally defined by polyclonal antisera raised for the purpose, but more recently monoclonal antibody reagents have also sometimes been used. As technology has developed, a large number of molecular techniques have been used to discriminate and group isolates with variable levels of success (Achtman, 1996).

Resolving whether two isolates have a common origin is essentially a genealogical question of how closely those isolates are related, establishing when and where their last most-recent common ancestor (MRCA) existed. For example, if two identical food-borne pathogen isolates from different patients share the same source, then the MRCA of those two pathogens will have existed in that source (van Belkum et al., 2001). In some cases, such antigens may be encoded by phase variable genes, so very closely related isolates that share a recent common ancestor may 'type' as being quite different. On the other hand, all members of the same species will share a common ancestor. Therefore, efficient molecular typing is a matter of choosing the characterisation method that achieves the appropriate level of resolution for the particular epidemiological question in hand or application envisaged (Maiden et al., 2013).

The advent of molecular techniques, and especially sequence-based techniques, provides a wealth of data from which genealogical relationships can be deduced and it is increasingly possible to design typing approaches that specifically address the question of common ancestry at a particular level of resolution (Maiden et al., 2013). Typing schemes can also be devised that reflect these genealogical relationships; indeed, it is now well established that *Salmonella* serotypes, for example, were successful as epidemiological indicators because the antigen-encoding loci predicted relationship by descent reasonably accurately so that the typing using antisera reagents can be entirely replaced with molecular methods (Achtman et al., 2012). The diversity of bacteria presents challenges: different pathogen populations exhibit different levels of diversity, are structured to different extents, and evolve at different rates, confusing the determination of the time of existence of the MRCA. Consequently, even with very high-resolution molecular approaches, up to and including whole genome sequence (WGS) analyses, it is not possible to establish whether two isolates are epidemiologically related without an appreciation of the population diversity and structure of the bacterium in question (van Belkum et al., 2001).

Knowledge of bacterial population structure and the forces that generate and maintain it are critical to effective molecular epidemiology for bacterial pathogens. This section will describe the basic principles of the subject with specific comments on the population biology of the genera *Campylobacter, Salmonella, Escherichia* and *Listeria*.

2.1.2. Forces that structure bacterial populations

Bacteria have been present on the earth for around 3.5 billion years, and have evolved to occupy all possible biological niches. Consequently, bacterial populations are both very large and have accumulated very high levels of diversity. As bacteria live in direct contact with their environment, they have evolved multiple mechanisms to mediate phenotypic variation, including complex genetic

switches or 'contingency genes' (Moxon et al., 1994) with the ability to transfer deoxyribonucleic acid (DNA), often among very distantly related organisms (Lorenz and Wackernagel, 1994). Organisms which cause disease when growing in multi-cellular hosts have evolved on multiple occasions, and represent subsets of this diversity. In contrast, individual pathogen populations, not least those of the genus *Campylobacter*, *Escherichia*, *Salmonella* and *Listeria*, have extremely large population sizes and extensive diversity, with continual mutational changes which facilitate evolution of the bacterial population (Wilson et al., 2009). It can be estimated, for example, that on the global scale each base pair in the *E. coli* genome is mutating multiple times each day. This makes bacterial populations potentially very responsive to changes in the environments by natural selection, as has been demonstrated by the rise of antibiotic resistance. It is worth noting that this is not necessarily mediated by mutation; natural selection can also lead to the increase in prevalence of previously existing rarer variants, which can be indistinguishable from the appearance of new mutant variants.

All of the organisms discussed in this Opinion have multiple mechanisms to evade the host immune response to maximise their residency and potential to reproduce within the host. For continued spread, all infectious agents must have a basic reproductive number (R0) greater than or equal to 1 (Anderson and May, 1979), this can be achieved by various combinations of duration of infectiousness of the host and likelihood of transmission. In general terms, acute, short-term, infections, require high rates of transmission per unit time, while chronic, or long-term infections can be successfully transmitted at lower intensity. Acute infections are therefore often more symptomatic than chronic infections; organisms causing diarrhoea representing just one example. Apparently very similar organisms may have evolved alternative strategies and these are sometimes influenced by episomal elements such as phage, plasmids and transposable elements (Hacker and Carniel, 2001). The continual opportunity for microbial evolution offers challenges to control but also provides a means of identification of epidemic and emerging strains and those involved in point source outbreaks by means of evaluation of genetic changes. Early methods of identifying genetic differences among bacteria, such as plasmid profile typing, assess only a small part of the genome that may be subject to change over a relatively short time, but newer techniques can identify genes that are more likely to be directly related to the functionality of the organism and its place in the general evolution of the species. The main challenge is to interpret the large quantities of data that are being made available in a meaningful way, using harmonised methods that give equivalent results regardless of the instrumentation or operator (Maiden et al., 2013).

While the bacterial population biology remains incompletely understood, substantial advances have been made in describing the population structuring of bacterial species (Achtman and Wagner, 2008), including the four food-borne pathogens which are the subject of this Opinion, and some progress has been made in understanding the mechanisms whereby these structures arise and are maintained.

A major consideration is the inherently clonal nature of bacterial populations, which is a consequence of their asexual reproduction acting in combination with diversity reduction events such as periodic selection and bottlenecking (Figure 3). Where these are the only, or predominant, mechanisms operating during bacterial population growth and spread, they inevitably result in a clonal population structure. This exhibits a tree-like phylogeny with the properties of congruence (i.e. the same phylogenetic signal is recorded at all loci) and linkage disequilibrium (the non-random assortment of genetic variation at multiple loci) (Gupta and Maiden, 2001). Such population structures are, in principle, simple to understand. Bacteria that conform to these expectations are relatively straightforward to study as gathering sufficient resolution of sequence variation will enable easy and quick determination of relationships among isolates. Some highly uniform 'single clone pathogens', for example *S. enterica* var. Typhi, *S.* Enteritidis PT4 and *S.* Typhimurium DT 104 have been difficult to study for the purposes of molecular typing since they have highly clonal population structure, due to very low levels of diversity, but these problems can now be successfully addressed with WGS approaches (Lan et al., 2009; Tankouo-Sandjong et al., 2012).

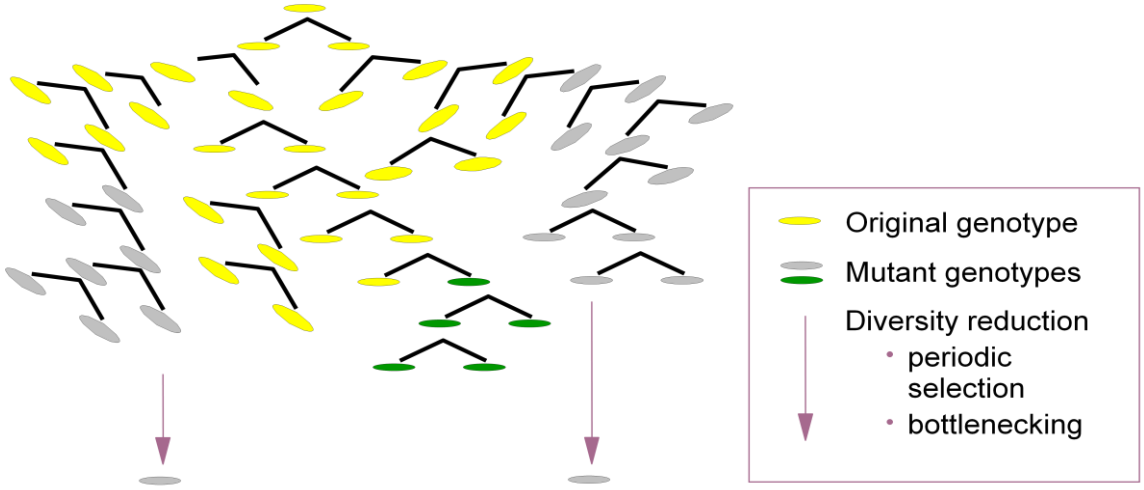

Figure 3: Clonal population structures in bacteria. All bacteria reproduce asexually by binary fission. If this is the only process whereby genetic information can be transmitted (that is vertically from parent to offspring) mutations only occur in the descendants of those organisms in which they occurred. Combined with diversity reduction events, which can be caused by selection for a particular characteristic (e.g. resistance to a given antimicrobial) or chance, this leads to a clonal population structure comprising lineages of bacteria.

In practice, the majority of bacterial populations do not conform strictly to clonal models, as they exhibit variable rates of horizontal gene transfer (Figure 4) (Gupta and Maiden, 2001). This process disrupts clonal structure due to descent, sometimes completely and has major implications for molecular epidemiological studies, as the properties of tree-like population structure, congruence and linkage equilibrium are all distorted (Maynard Smith et al., 1993). This means that the uncritical application of tree reconstruction methods and inappropriate sampling of genetic variation can distort the determination of relationships among isolates. Evidence accumulated to date demonstrates that rates of recombination differ among different bacteria with corresponding differences in the levels of disruption of clonal structure (Didelot and Maiden, 2010). For most bacteria, disruption of clonal structure has occurred and therefore analysis approaches have to take this into account if the data collected are to accurately reflect the biology of the organism in question.

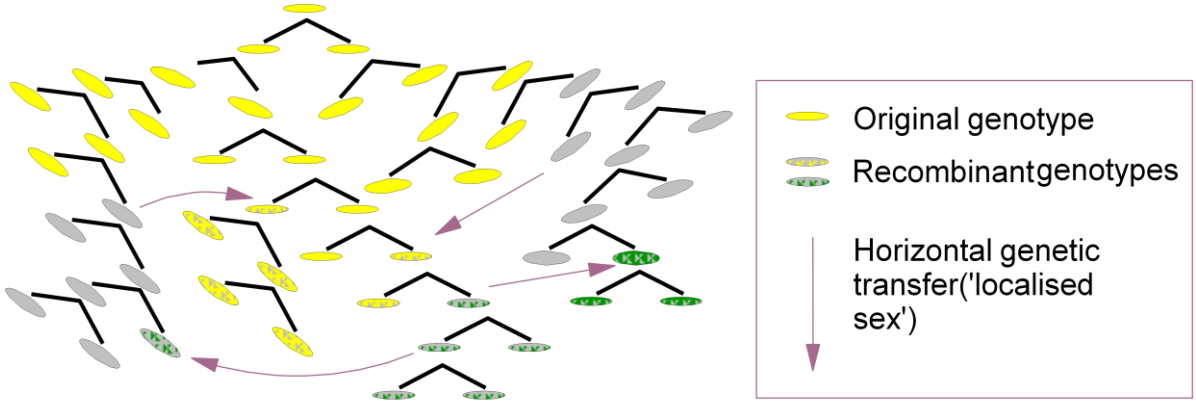

Figure 4: Effect of horizontal genetic transfer on bacterial population structure. In the presence of mechanisms of horizontal genetic exchanges (that is, transfer of genetic material among bacteria that do not necessarily share a parent) genes are mobilised widely and less likely to be lost from the population by diversity reduction events.

To be successful, typing techniques should reflect both the genealogical and the functional relationships of bacterial isolates. These can be complicated by the presence of accessory elements of

the genome, such as phage and plasmids, which are frequently highly mobile among lineages which are defined by sequence variation in the core genome; those genetic elements that are universally present in a group of isolates. In principle, typing of clonal organisms (Figure 3) is usually straight-forward as treelike phylogenies accurately represent genealogical relationships among isolates and can be readily reconstructed from genetic data. The lack of horizontal genetic exchange will also mean that once a lineage has acquired a particular characteristic (e.g. antigen) it is likely to retain it. Recombining, non- or semi- clonal organisms are more challenging for epidemiology as they are less likely to be uniquely associated with particular characteristics. In practice, this usually means that for these organisms it is necessary to sample genetic variation from multiple locations in the chromosome, preferably including parts of the genome for which there is at least some information concerning their evolution. The extent to which a particular genetic or phenotypic marker is reliable as an indicator for both genetic and functional relationships of isolates determines its applicability in particular bacterial species. The biological basis of this relationship does not have to be known, but it is highly preferable that this is so. For example, for some *Salmonella* serovars pulsed gel electrophoresis (PFGE) profiles have proved to be valuable, although by no means perfect, indicators of genetic and phenotypic relationships among different lineages. On the other hand, *Campylobacter* serotypes provided limited information on their population structure as these types were phase-variable and extensively re-assorted among genetically diverse bacteria by horizontal genetic transfer and PFGE is only suitable for short-term investigations of outbreaks of *Campylobacter* for the same reasons.

2.2. *Campylobacter* spp.

Campylobacteriosis is the most prevalent cause of human bacterial gastroenteritis worldwide. The genus *Campylobacter* comprises a number of species several of which have been implicated in human and animal disease, but by far the most important of these in terms of food-borne disease are the 'thermophilic' species *Campylobacter jejuni* and *C. coli*. The precise contribution of these two species to human diseased burden varies with setting and over time, but in countries such as the UK, for example, approximately 90% of cases are caused by *C. jejuni* and 10% by *C. coli*. These organisms are widespread in farm and wild animals and can be isolated from many environmental reservoirs. They are common contaminants of retail food, especially broiler meat, and it is now established that the latter is a major source of human disease in most countries, although infection from other food and environmental routes also occurs (EFSA Panel on Biological Hazards, 2011a).

Campylobacter jejuni and *C. coli* are closely related but distinct species which share about 87% nucleotide sequence identity across their genomes. Consequently, they can be difficult to distinguish, in large part because of widespread recent introgression of *C. jejuni* genes into some *C. coli* lineages. Intriguingly, whilst sharing many properties, these two species have distinct population structures. Three distinct clades have been described in *C. coli*, with limited evidence for recombination among them. Introgression of *C. jejuni* genes appears to be limited to clade 1, which comprises organisms mostly associated with poultry and human disease. Clades 2 and 3 are mostly found in wild animals and environmental sources, are not introgressed with *C. jejuni* genes, and are rarely associated with human disease (Sheppard et al., 2008; Sheppard et al., 2011a; Sheppard et al., 2013a). By contrast, *C. jejuni* comprises a large number of genetic types, with widespread evidence for recombination throughout the species. Notwithstanding recombination, the extensive diversity of these organisms is structured into a large number of distinct lineages identified as 'clonal complexes' by multilocus sequence typing (MLST), which is also a reliable approach for the speciation and the identification of *C. coli* lineages (Colles and Maiden, 2012).

The identification of *Campylobacter* clonal complexes proved to be a major advance in understanding *Campylobacter* epidemiology, enabling the reproducible grouping of isolates for the first time (Dingle et al., 2002). Further, it was rapidly established that given clonal complexes were associated with specific hosts. Many *Campylobacter* clonal complexes are associated with particular host animals (Sheppard et al., 2011b). This has been especially studied in wild birds, where given *Campylobacter* genotypes are associated with particular host species, a pattern which is robust to geographic variation

at global distances (e.g. the *Campylobacter* isolated from a particular bird species is the same whether the isolates have been obtained in Europe or Australasia) (Griekspoor et al., 2013).

Of particular interest to the investigation of food-borne disease, is the fact that there are a number of *Campylobacter* genotypes, as defined by clonal complex and including both *C. jejuni* and *C. coli*, which are associated with farm animals including those from poultry, bovine, ovine and porcine sources. Knowledge of these associations has enabled genetic attribution studies that investigate the most likely sources of human infection, indicating that the majority of human campylobacteriosis in many industrialised countries is likely to come from agricultural, predominantly, chicken sources (Wilson et al., 2008; Mullner et al., 2009; Sheppard et al., 2009). Intriguingly, given the differences normally seen among different wild bird species, some of the major human disease-associated *C. jejuni* clonal complexes, especially the ST-21 complex, occur in chicken and ruminant (specifically bovine and ovine) agricultural sources (McCarthy et al., 2007). These 'multi-host' genotypes are a major factor in genetic attribution studies, although the prevalence of *Campylobacter* in retail chickens in many high-income countries makes this the most likely source for most human infection. In conclusion, the population structures of *C. jejuni* and *C. coli* have been well-described along with the impact that this has on human disease and these approaches have already had major impacts on the development of control methods for these organisms.

Although human campylobacteriosis is very widespread, point source outbreaks that can be identified and controlled by public health action are rare. Where they do occur, such outbreaks are often associated with poor food preparation at particular institutions, unpasteurised milk or milk pasteurisation failures (Taylor et al., 2012) or contaminated water (Craun et al., 2010). In the case of food preparation related outbreaks, the genetic diversity of the outbreak-causing bacteria can be high, if the source material, usually of poultry origin, contains multiple genotypes.

2.3. *Salmonella* spp.

Salmonellosis, manifested mainly by enteric fever and gastro-enteritis, is the second most commonly reported bacterial cause of food-borne illness in the EU. The organism is widespread as a normally subclinical infection in food production animals, which are the major source of zoonotic infection.

The genus *Salmonella* comprises two species: *enterica* and *bongeri*. Within *enterica* there are seven subspecies of which subspecies *enterica* contributes most animal and human pathogenic strains. The other subspecies, as well as *S. bongeri*, are more likely to be linked with reptiles and environmental niches. Although the subspecies were first defined by antigenic and biochemical differences, the subspecies are quite distinct in terms of DNA sequence variation (Lan et al., 2009). The frequency of genetic recombination varies dramatically between the subspecies, with subspecies *enterica* showing the greatest diversity and a low level of clonality (Octavia and Lan, 2006). Didelot et al., (2011) identified five distinct lineages within this subspecies, with evidence of the highest levels of recombination being within lineages.

Within subspecies there are also large differences, with *S.* Typhimurium, considered to be an ancient lineage, showing the greatest variability and *S.* Typhi, a human host-adapted serovar, being of more recent origin and showing more limited diversity (Lan et al., 2009). *S.* Newport provides an example of the existence of at least 3 distinct lineages within a single serovar (Sangal et al., 2010), with wildlife, cattle and turkeys as predominant hosts in Europe. The recently emerged multi-drug resistant (MDR) AmpC-producing strains are represented by two MLST sequence types (STs) within one of the lineages. Four lineages were found amongst isolates of *S.* Paratyphi B, including variant Java (Sangal et al., 2010). The development of different clonal groups of *Salmonella* is largely related to expansion of subtypes within serovars (e.g. *S.* Typhimurium DT 104) in response to epidemiological opportunities presented by changes in the immune status of animal and human populations or management practices (Desai et al., 2013). The concept of serovars is likely to be redefined by molecular studies which show substantial genetic diversity within serovars such as Newport (Cao et al, 2013), but there are also examples of situations in which genetic studies have led to consolidation of

previous distinct serovars, such as *S.* Gallinarum and *S.* Pullorum, which have very different clinical and epidemiological properties (Tang et al, 2013).

Horizontal gene transfer of antimicrobial resistance genes, sometimes with linked virulence genes, can be associated with the emergence of new groups of strains with 'epidemic' potential (Alcaine et al., 2005). Lateral gene transfer is thought to have contributed up to 25% of the gene repertoire of some *S.* Typhimurium strains (Porwollik and McClelland, 2003). *Salmonella* bacteriophages are often involved and mediate horizontal transfer of virulence genes among *Salmonella* strains by transduction and lysogenic conversion (Rabsch et al., 2002b). Integrons, genomic islands and plasmids (which may carry integrons with them) may be involved in transfer of bacterial secretion systems, virulence and antimicrobial resistance genes (Guerra et al., 2001; Dobrindt et al., 2004; Randall et al., 2004; Sandvang et al., 2006, Bhatty et al., 2013) from other organisms to *Salmonella*, and *vice versa* (Siebor and Neuwirth, 2013).

The reduction of major *Salmonella* serovars has sometimes been followed by the emergence of others and it is postulated that a '*Salmonella* niche' may exist in animal production that is more likely to be filled by emergent strains than if the original strains persisted (Rabsch et al., 2000). This is specifically suspected in relation to the elimination of *S.* Gallinarum biovar Pullorum (SP) from commercial scale poultry production in many countries, which may have provided an immunologically naïve niche for the emergence of *S.* Enteritidis.. The evolution of virulence in *Salmonella* is largely driven by horizontal gene transfer and this has given rise to highly flexible pathogens that are able to colonize new niches and extend their host range (Bäumler, 1997).

The number of cases of *S.* Typhimurium in cases of human infection in Europe has reduced gradually over time, despite the emergence of monophasic strains, largely in parallel with the unexplained decline in *S.* Typhimurium definitive phage type (DT) 104 (EFSA and ECDC, 2012). This contrasts with the dramatic reduction in *S.* Enteritidis that was made possible by strict legislative control measures in chicken breeding and commercial laying hen flocks across Europe. It is likely that the periodic emergence, international dissemination, predominance and decline of certain *S.* Typhimurium phage types/clonal groups is related to an ability to initially evade protective immune responses associated with prior exposure to unrelated *S.* Typhimurium strains, occurrence in breeding animals that are traded internationally and presence of multiple resistance to commonly used antimicrobials which may aid their selection in medicated animals (Rabsch et al., 2002). Regression of such strains may be due to development of herd immunity across animal populations, reduction in virulence or persistence ability amongst strains and their replacement by other strains.

The emergence of monophasic strains of *S.* Typhimurium with the antigenic formula 1,4,(5),12:i:- was a further development, firstly in Spain with U302 strains and from 2006 in most of the rest of the EU, this time involving DT 193/120 isolates with a characteristic tetra resistance pattern of ASSuT (A, ampicillin, S, streptomycin, Su, sulphonamides, T, tetracyclines) (EFSA Panel on Biological Hazards, 2010). In some cases resistance genes may be linked with virulence genes on the same transferable genetic elements. In the United States of America (USA), different monophasic *S.* Typhimurium strains have emerged over a similar time scale and the reason for this concurrent emergence of multiple clones is not known (Bugarel et al., 2012). Loss of the phase 2 flagella antigens has been postulated as a mechanism by which organisms may partially evade the initial cytokine response in host animals (Crayford et al., 2011). A further shift involving an increase in the proportion of monophasic *S.* Typhimurium strains lacking the O5 antigen is also thought to provide similar immunological advantages to the organism (Slauch et al., 1995). The timing of the emergence of monophasic *S.* Typhimurium corresponds with the withdrawal of antibiotic growth promoters and increasing use of zinc oxide supplements in feed to help suppress bacterial overgrowth in the intestines of weaned pigs. Monophasic *S.* Typhimurium DT193 isolates have two genomic islands, one of which codes for both the tetra-resistance and resistance to heavy metals. Use of heavy metals in pigs may therefore have been involved in the preferential selection and proliferation of monophasic *S.* Typhimurium mutants (Gebreyes, 2011).

2.4. *Escherichia coli*

2.4.1. The concepts of STEC, EHEC and EAEC

Shiga toxin-producing *Escherichia coli* (STEC) are characterised by the production of Shiga toxins (Stx), also known as Vero toxins (Vtx) because of cytotoxicity to Vero cells. Illnesses associated with STEC, also known as verotoxigenic *E. coli*, verocytotoxigenic *E. coli*, and verotoxin-producing *E. coli* (VTEC), range from mild to bloody diarrhoea to haemorrhagic colitis (HC), haemolytic uraemic syndrome (HUS), and thrombocytopenia. Such symptoms are common to STEC infections worldwide and are particularly serious in children. The most important serotype worldwide that is associated with both outbreaks and sporadic cases has been *E. coli* O157:H7, although an increasing number of non-O157 STEC infections have been recorded in the EU over recent years. The most common non-O157 serotypes from cases of human infection in the EU from 2007 to 2010 were O26:H11, O103:H2, O145:H-, O191:H-, O111:H-, O117:H7 and O146:H21, although infections with at least ten further serotypes were notified (EFSA BIOHAZ Panel, 2013). Many other non O157 STEC serotypes have been identified in food animals and derived foods but their capacity for causing severe disease in humans is as yet unknown (EFSA BIOHAZ Panel, 2013). Recent developments, and in particular the increasing numbers of reports of non-O157 STEC outbreaks, including the major outbreak of serotype O104:H4 in 2011 (see below), has focused attention on the risk associated with strains of STEC other than O157:H7. These are less easily identified, particularly in food-related sources and in animal or environmental reservoirs.

EHEC were originally defined as a subset of STEC, that were associated with haemorrhagic colitis, and later as certain O groups (O26, O45, O103, O111, O121, O145 and O157) that in addition to the *vtx*-encoding genes, usually carry the attaching and effacing gene (*eae*; intimin-coding) and thereby have the ability to cause attaching and effacing (A/E) lesions in infected cells. The ability to cause A/E lesions is mediated by the locus of enterocyte effacement (LEE) pathogenicity island (PAI). EHEC strains are typically isolated from cases of severe disease, but are poorly defined because there is no commonly accepted definition of EHEC. In some countries EHEC is synonymous with VTEC.

Enteroaggregative *E. coli* (EAEC) are characterised by their ability to aggregatively adhere to tissue culture cells in a distinct 'stacked and brick-like' manner which is mediated by aggregative adherence fimbriae (AAF). They usually produce an enteroaggregative heat-stable toxin (EAST1) encoded by plasmid-borne *astA* genes.

Within STEC, EHEC and EAEC, isolates belonging to specific STEC pathotypes and one EAEC pathotype (EAEC O104:H4) are of particular importance in the context of food safety.

2.4.2. Virulence attributes of STEC and EAEC

There is no single marker or combination of marker(s) that defines the potential of a STEC strain to cause human disease, as various factors and toxins contribute to the virulence of STEC. Shiga toxin type 2 (Stx2) is more often associated with confirmed cases of human disease, and those strains producing this toxin are more frequently associated with severe illness. Strains that produce Stx2 have been suggested to be more likely to cause HUS than those that produce Stx1 alone (Bosilevac and Koohmaraie, 2011; Friedrich et al., 2002). *Stx* gene sub-typing can also provide an indication of risk, with certain subtypes being more likely to result in serious illness (Scheutz et al., 2012; Soborg et al., 2013). Intimin genes can also be sub-typed and related to severity of disease (Wang et al., 2013).The presence of *stx*1 and *stx*2 genes has recently been recorded in the annual European Union (EU) surveillance report (EFSA and ECDC, 2013).

Except for the intimin protein encoded by the *eaeA* gene within the LEE PAI and the AAF encoded by the EAEC PAA plasmid and regulated by the transcriptional AggR regulon, no other adherence factors have been consistently associated with the virulence of STEC and EAEC, respectively (Nataro and Kaper, 1998; Nataro et al., 1998; Kaper et al., 2004), although genes associated with quorum sensing, such as *Sdi*A, have been shown to be essential for efficient colonisation of carrier animals (Sharma

and Bearson, 2013b). Patient-associated (e.g., age, immune status, antibiotic therapy in the pre-infection period), and dose-related factors are also important.

Detection of specific genes can also be used as a method for screening and presumptive identification of STEC in samples from primary production or food. ISO/TR 13136:2012 describes the identification of STEC by means of the detection of the following genes: (a) the major virulence genes of STEC, *stx* and *eae*; (b) the genes associated with the serogroups O157, O111, O26, O103, and O145. When one or both of the *stx* genes is detected, the isolation of the strain from the pre-enriched sample is attempted.

2.4.3. The 2011 *E. coli* O104:H4 outbreak

The most recent example of a major outbreak caused by a non-O157 STEC was the O104:H4 outbreak, first identified in northern Germany in May 2011 (Frank et al., 2011a). This outbreak resulted in 4 321 confirmed of STEC infection and 852 of HUS, with 54 deaths reported in 14 EU countries, the USA and Canada when the epidemic was declared to be over at the end of July 2011(Buchholz et al., 2011; Karch et al., 2012). The outbreak was unusual because of the high proportion of adult patients (ca. 25 %) presenting with HUS, plus the frequent development of neurological symptoms in these patients (Frank et al., 2011b). These clinical characteristics were thought to be due to the unique combination of traits carried by the pathogen, which included features typical of EAEC, together with the capacity to produce Stx (Frank et al., 2011b). The strain also has a distinct set of additional virulence and antibiotic resistance genes, including genes encoding resistance to a range of cephalosporin antibiotics (Rasko et al., 2011). Whole genome sequence analysis has suggested that the clinical characteristics of the outbreak strain were due to the unique combination of virulence factors carried by the pathogen and acquired by horizontal gene transfer (Frank et al., 2011a, b).

2.5. *Listeria monocytogenes*

In the EU, listeriosis is a relatively rare but nevertheless serious food-borne illness in humans, with high morbidity, hospitalisation and mortality in vulnerable populations. *Listeria* species are ubiquitous organisms that are widely distributed in the environment, especially in plant matter and soil. The principal reservoirs of *Listeria* are soil, forage and surface water. The main route of transmission to humans is believed to be through consumption of contaminated food. The bacterium can be found in raw foods and in processed foods that are contaminated during and/or after processing.

The population structure of *Listeria* is highly clonal and can be divided into four major lineages which are correlated with serotype groupings. Within the four lineages of *L. monocytogenes*, strains are generally classified by serotyping or MLST of which serotypes 1/2a, 1/2b and 4b are most commonly (98% cases) associated with human listerial infections and serotypes 4a and 4c are rare; serotype 4b is more likely to associated with outbreaks than sporadic cases whereas lineages III and IV are relatively rare (Wiedmann et al., 1996).

The genus *Listeria* currently comprises 10 species of Gram-positive, facultatively anaerobic, non-spore forming bacteria (*L. fleischmannii, L. grayi, L. innocua, L. ivanovii, L. marthii, L. monocytogenes, L. rocourtiae, L. seeligeri, L. weihenstephanensis* and *L. welshimeri*), but cases of human listeriosis are almost exclusively caused by the species *Listeria monocytogenes* (EFSA and ECDC, 2013).

The diversity and ecology of *Listeria* spp. has been summarised by den Bakker et al. (2010). *L. monocytogenes* and *L. ivanovii* are pathogens of warm-blooded hosts. *L. monocytogenes* causes severe food-borne disease in humans, with a mortality rate of approximately 20–30%, as well as invasive infections in a number of other warm-blooded host species, particularly ruminants. *L. ivanovii* predominantly causes infections in ruminants, but has also been associated with rare infections in humans.

Listeria monocytogenes can persist within food processing environments for long periods of time, due in part to its ability to grow at wide-ranging temperatures and pH (-0.4°C to 45°C, optimum 37°C; pH 4.39 to 9.4, optimum 7.0 (ICMSF, 1996)) and the ability to form biofilms promoting adherence to food processing surfaces. The persistence of a single subtype of *L. monocytogenes* in processing facilities or on equipment has been reported for up to 10 years. These characteristics are fundamental to the routes of contamination of specific foodstuffs that are commonly involved in infection.

The two pathogenic *Listeria* species are closely related to non-pathogenic species; *L. monocytogenes* is closely related to *L. innocua* and *L. marthii*, and *L. ivanovii* is closely related to *L. seeligeri*, which is non-pathogenic even though many isolates contain a homologue of the main *Listeria* virulence gene cluster. The main *Listeria* virulence gene cluster (also known as the *prfA* virulence cluster or the *Listeria* pathogenicity island [LiPI]) encodes a number of proteins that are necessary for intracellular survival and motility. Specific functions encoded in this cluster include hemolysin, two phospholipases and a metalloprotease (encoded by *hly*, *plcA*, *plcB*, and *mpl*), which all contribute to escape from host cell vacuoles, an actin polymerizing protein (encoded by *actA*), and a global regulator of virulence gene transcription (encoded by *prfA*). Members of the internalin protein family, which are cell wall anchored or secreted proteins that are characterized by the presence of leucine rich repeats, are also associated with virulence in various *Listeria* strains. While a considerable number of genes encoding internalins have been found in pathogenic and non-pathogenic *Listeria*, clear virulence related functions have only been assigned to a few internalins, including *inlA* and *inlB*, which encode proteins required for invasion of different cells types, including human intestinal epithelial cells, and *inlC*.

Genome evolution in *Listeria* has involved limited gene loss and acquisition as suggested by relatively high coverage of the predicted pan-genome by the observed pan-genome, conserved genome size (between 2.8 and 3.2 million of base pairs) and a highly syntenically linked genome. Limited gene loss in *Listeria* includes loss of virulence associated genes, likely associated with multiple transitions to a saprotrophic lifestyle. The genus *Listeria* thus provides an example of a group of bacteria that appears to evolve through a loss of virulence rather than acquisition of virulence characteristics. While *Listeria* includes a number of species-like clades, many of these putative species include clades or strains with atypical virulence associated characteristics (den Bakker et al., 2010).

2.6. Concluding remarks on relationships of pathogen population structure and epidemiology to molecular epidemiology

Transmission of DNA among bacterial species by means of mechanisms such as phage or plasmids is common to most bacterial species and can result in acquisition of virulence, resistance or stress response genes that enhance the survival of organisms in different host and environmental niches. These genetic characteristics are fundamental to the epidemiology of the organisms, but these elements has until now been difficult and laborious to in characterise. Modern molecular epidemiology will be equipped with tools to fully elucidate the role of the genomic elements in pathogens and to use this information for tracking strains and assessing their likely significance in the food chain.

Campylobacter jejuni is an extremely diverse bacterium with a fundamentally non-clonal population structure with extensive evidence for horizontal genetic exchange; however, the use of multilocus techniques has identified clonal complexes, groups of organisms that share a common ancestor. These clonal complexes are valuable units of analysis, as they are associated with characteristics of interest such as host association and survival in the environments and the human food chain. The related species, *C. coli*, is divided into three lineages or 'clades', only one of which is regularly associated with human disease. This clade has been substantially introgressed on at least two separate occasions with *C. jejuni* genes.

Salmonella is a very large and diverse bacterial genus that has adapted and diversified to occupy a wide variety of host and environmental niches, often involving specific host adaptation. Two serovars (*S.* Enteritidis and *S.* Typhimurium) dominate human infection in most countries and their success

appears to be associated with epidemiological relationships with certain food animal species, providing the means for widespread exposure. The evolution of *Salmonella* is largely driven by horizontal gene transfer and this has given rise to highly flexible pathogens that are able to colonize new niches and extend their host range. This may also be linked to the emergence of various epidemic strains of *Salmonella* in combination with other biological factors and epidemiological opportunities for dissemination.

The complexity of STEC relates to the difficulty of designating individual strains as pathogens, since there are many apparently non-pathogenic strains with *stx* or other genes that might be expected to be more virulent in humans. The plasticity of the genome, resulting in the acquisition of virulence or adherence properties from other organisms, normally by means of translocation on phages means that new and unexpected strains are likely to appear in an unpredictable way over time. The driving force for the emergence of STEC in ruminant populations is not understood and intervention measures are also uncertain, although vaccination and certain management practices may help reduce carriage of these organisms. The major 2011outbreak of *E. coli* O104:H4 is an example of the genomic variability referred to above and has seriously challenged the concept of STEC seropathogenicity (Karmali et al., 2003). This outbreak has demonstrated that horizontal gene transfer can result in the formation of a new, highly pathogenic strain with the virulence characteristics of both STEC and EAEC strains, despite lacking *eae* genes that were thought to be essential for the expression of virulence in 'traditional' STEC. This outbreak has also demonstrated the value of WGS in understanding the phylogeny of the causative strain. Although in this instance such studies were not essential for disease control, they have clearly shown the value of NGS techniques for understanding the evolution of virulence and pathogenicity in common food-borne bacterial pathogens.

Listeria monocytogenes is an intracellular bacterium responsible for a disease characterized by diverse clinical presentations that are linked with genetic change associated with different lineages and serotypes. The population structure of *Listeria* is highly clonal and can be divided into four major lineages which are correlated with serotype groupings. Genome evolution in *Listeria* has involved limited gene loss and acquisition as suggested by relatively high coverage of the predicted pan-genome by the observed pan-genome, conserved genome size (between 2.8 and 3.2 million base pairs) and a highly syntenically linked genome. Limited gene loss in *Listeria* includes loss of virulence associated genes, likely associated with multiple transitions to a saprotrophic lifestyle.

Clinical features are attributed to its unusual capacity to cross several barriers in susceptible hosts. Intracellular infection is a consequence of the bacterium's capacity to enter and replicate in a wide variety of mammalian cells, and to its faculty to spread from one cell to the next, thereby escaping the humeral immune response.

In summary, the following are concluding remarks on the structure and epidemiology of bacterial populations:

- All bacteria are subject to genetic change (e.g. in response to environmental stress and human interventions such as antimicrobial or heavy metal use or vaccination), sometimes by mutation but more often by acquisition or loss of genetic elements. These changes can be followed by clonal expansion in the case of biologically successful organisms.

- Ongoing evolution driven by genetic mutation and selection has given rise to highly adaptable organisms that are able to exploit and expand into novel niches and extend their host range. Such evolution may also be linked to the emergence of various 'epidemic' strains of pathogens, such as *Salmonella*, in combination with other biological factors and epidemiological opportunities for dissemination.

- Typing methods can be devised that measure variation, in the core genome (genes that are invariably present in a particular group of isolates) and the accessory genome (genes variably

present in the genome). The accessory genome includes plasmids and phages, which although independent of the bacterial host genome are frequently involved in virulence.

- The molecular characteristics of organisms provide markers for investigation of outbreaks, attribution studies, and assessment of potential virulence or epidemic potential.

- Even with high-resolution molecular approaches, up to and including whole genome sequencing (WGS) analysis, it is not possible to establish how closely two isolates are related without an appreciation of the structure and diversity of the bacterial population in question.

3. Review of current and prospective molecular identification and typing methods

Genetic methods for bacterial typing have progressively replaced phenotypic assays during the last two decades, even though the phenotypic methods are still widely used by reference laboratories for routine surveillance and outbreak detection, as reported in an EU-wide survey (EFSA, 2009). The current practice is to use a combination of different phenotypic and genotypic typing methods.

During the last three decades, a large number of genotyping methods have been developed and applied in various contexts, mostly by research institutions or reference laboratories dealing with local or national outbreaks. Difficulties in standardisation and harmonisation of the results have often made data difficult to share. For some methods standardisation and harmonisation has been developed to a degree that has made application of the methods suitable for wider international use (e.g. Pulsenet International[8]).

Certain typing methods can provide information on the phylogenetic relationships between organisms. This information has potentially more epidemiological value than the identification of random markers that indicate genetic differences that cannot be linked with phylogeny and function (Hazen et al., 2013; Wilson et al, 2012). Thus, methods that can act as indicators of evolutionary processes, such as MLST and single nucleotide polymorphism typing (SNP) and that can be linked with phylogeny as well as profiling of virulence genes, genomic islands and other relevant genes, can be more informative than less specific typing methods such as PFGE and Restriction Fragment Length Polymorphism (RFLP). WGS can, in principle, be used to derive any desired combination of genetic targets to provide maximum epidemiological relevance - according to the interpretation pipelines used - and thus are likely to replace the combinations of other techniques that are in current use as cost and complexity reduces and familiarity increases (Chen et al., 2013; Sabat et al., 2013). Nevertheless, molecular typing analysis can also be useful independently of phylogenetic analysis, as such methods may identify the relatedness of individual strains. In some cases methods may include the identification of certain markers that are important for interpreting phenotypic characteristics relevant for public health, e.g. antimicrobial resistance or virulence genes.

The focus in this section is on the general principles of main molecular typing methods that are currently used and prospective methods for epidemiological typing of the main food-borne bacteria (*Campylobacter, Salmonella*, STEC and *Listeria*) by national and international reference laboratories. For more comprehensive information on typing methods, refer to recent reviews (e.g. Li et al., 2009, Hallin et al., 2012; Sabat et al., 2013; MacCannell, 2013). This section will present and discuss methods that are widely used in many laboratories, even though it is acknowledged that a range of other methods exist which are not widely used internationally (e.g. typing of clustered regularly interspaced short palindromic repeats (CRISPR) loci, octamer-based genome scanning (OBGS), comparative genomic hybridization (CGH)).

A set of pre-defined criteria based on the first term of reference of the mandate was established for evaluating the individual molecular typing methods. These criteria and its interpretation are as follows:

1. **Discriminatory capability**: What is the degree of discrimination between strains of different genotype? The 'discriminatory power' of a typing method is defined as the average probability that the method will assign a different type to two unrelated strains randomly sampled in the microbial population of a given taxon (Hunter, 1990).

2. **Reproducibility and repeatability**: What is the consistency of results within and between laboratories, and over time;

3. **Current international harmonisation**: What is the current status with regards to the following parameters?

[8] Further information on PulseNet international available in Chapter 4 and at: http://www.pulsenetinternational.org/ (last visited on 11/12/2013)

a. Availability and use of Standard Operational Procedures (SOP);

b. Availability and use of External Quality Assurance systems (EQA);

c. Presence and use of harmonised nomenclature;

d. Availability and use of data management tools.

4. Potential for **future international harmonisation,** in situations where any of the above criteria (point 3) but which are not currently harmonised at international level may apply.

Further consideration should be given in the evaluation of the methods to practicalities such as skills, equipment needs and to the time taken to achieve meaningful results as well as total hands-on time and suitability for batching tests. The time component is also a practical issue that may be strongly linked to the specific operational capacity of the individual laboratories (i.e. between-laboratory variability may be expected). Thus, an absolute result for the time needed to perform each method and obtain meaningful results is not universal.

3.1. Current methods

3.1.1. Molecular serotyping

Molecular serotyping describes methods developed to identify serotypes of organisms by analysing DNA.

There are several ways that DNA-analysis can be used to achieve this. The most common methodology uses either one of these two key principles: (a) examination of the genetic loci known to produce the serologically reactive components used in traditional serotyping, or (b) examination of variations in the genome, which are indirectly associated with known serovars or serotypes. These variations may include various kinds of polymorphous regions, as long as they show a strong association to the traditional serovars/serotypes.

Several methods exist for *E. coli* and *Salmonella* where the O-antigen operon and the *fliC* gene have been sequenced and used directly for serotype determination, or specific Polymerase Chain Reaction (PCR) primers are constructed from the sequences to distinguish between conserved serotype determining genes. These are examples of principle (a) mentioned above. MLST used for serotyping *Salmonella enterica* subspecies *enterica* by examining variations in genome is an example of principle (b) (Achtman et al., 2012). Both these principles can be assayed using several different methodologies e.g. array technology (Guo et al., 2013), PCR (Paddock et al., 2012), real-time PCR (Anklam et al., 2012; Fratamico et al., 2011). For a more in depth study of molecular serotyping approaches for *Salmonella* see Ranieri et al., 2013; for *E. coli* see DebRoy et al., 2011; for *L. monocytogenes* see Doumith et al., 2005, Kérouanton et al., 2010 and Vitullo et al., 2013; and for *Campylobacter* see Poly et al., 2011.

For the four pathogens considered in this Opinion molecular serotyping is considered to provide a low to moderate *discriminatory capability*. This is normally similar or marginally higher than traditional serotyping as sub-types can often be recognised within serotypes. '*Reproducibility and repeatability'* are high, but may be reduced if large arrays are used, due to the complexity of the technology. **'***Internationally harmonised standards*' for molecular serotyping are not in place except for *L. monocytogenes*; nevertheless, the existing software tools could be employed at international level. Molecular serotyping is based on a well-known and implemented methodology, and thus has a high capability for **'***future international harmonisation'*. Molecular serotyping will, in most instances, provide results within a day from receiving the isolate.

3.1.2. Restriction Fragment Length Polymorphism (RFLP) analysis

In RFLP, a target DNA sequence known to show polymorphism between strains of the species of interest, is cleaved with restriction endonucleases to generate fragments of varying length. The earliest versions of the RFLP method involved several time-consuming steps. The whole process could in some cases take up to four weeks to produce an interpretable result.

In PCR-RFLP typing the target sequence is amplified at high annealing temperatures to maximise stringency. The amplified product is cut with restriction endonucleases and isolates are typed by comparing their RFLP pattern after gel electrophoresis. PCR-RFLP typing has provided limited discrimination. It can also be confounded either by mosaicism due to horizontal gene transfer (e.g., flagellin genes in *C. jejuni* (Harrington et al., 1997)), or by hypermutation at so–called contingency loci that undergo rapid rearrangements in response to environmental changes. When RFLP analysis is directed at genes encoding ribosomal ribonucleic acid (rRNA) the method is usually referred to as 'Ribotyping'. Ribotyping s has successfully been automated, and fully automated ribotyping is commonly referred to as 'riboprinting' after the RiboPrinter® commercial system (DuPont Qualicon, Wilmington, DE). Automated riboprinters require minimal input and technical skill by the operator, but the cost of equipment is high, so this method is largely used by commercial food companies.

Ribotyping has been reported to have limited resolution for *E. coli* O157:H7 (Martin et al., 1996), but a study from 2003 (Clark et al., 2003) concluded that ribotyping was useful in identifying clusters of *S.* Enteritidis, which is difficult to sub-type by single-enzyme PFGE. A study by Manfreda et al. (2003) ran multiplex PCR typing alongside riboprinting to demonstrate that the automated ribotyping method is highly reproducible and efficient enough to use as a library typing method for *Campylobacter* surveillance. For a review on the use of automated ribotyping in food safety see Pavlic and Griffiths (2009).

RFLP analysis may be regarded as providing a moderate to high *'discriminatory capability'* for at least some of the four pathogens of interest. Within and between laboratories '*reproducibility and repeatability'* of results is low to moderate for PCR-RLFP and traditional ribotyping, while high in the case of fully automated ripoprinting systems. At present, the riboprinting platform provided by DuPont Qualicon® seems to be the only RLFP typing that provides for '*internationally harmonised standards'*. Nevertheless, RLFP typing tools other than riboprinting also may have the '*potential for international harmonisation***'** in spite of the current lack of systems operating to achieve this. Automated riboprinters can deliver typing results within 24 hours of receiving isolates, whereas manual methods usually require considerably more time to completion.

3.1.3. Pulsed-Field Gel Electrophoresis (PFGE) analysis

PFGE was first described in 1984 and is currently the most frequently used DNA-based typing method for food-borne bacterial pathogens. The PFGE-method standardization and rigid quality control introduced by PulseNet International has resulted in PFGE becoming the most commonly used method for outbreak identification, surveillance and investigation for a number of important pathogens, in particular *Salmonella*, STEC and *Listeria* (Ribot et al., 2006). Thus, for these pathogens, the performance of new typing methods will be measured against PFGE.

PFGE is in effect a variant of the Restriction Endonuclease (REA) methodology. The abbreviation 'PFGE' does not describe a complete typing method but rather a technique to separate long strands of DNA through an agarose gel matrix and visualized as bands. The actual DNA content of each band is unknown and it may consist of one or multiple similar length fragments of different origins, thus in rare cases an identical match may be a result of homoplasy and not actual homology. The PFGE method has, proven immensely successful in many outbreaks, particularly of STEC and *Salmonella* (see Chapter 4).

PFGE fingerprinting has a high '*discriminatory capability*' for most pathogens considered, but with some notable exceptions e.g. *S.* Typhimurium DT 104, and *S.* Enteritidis PT 4, for which phenotypic

subtyping has already narrowed the selection and recent emergence has led to a high degree of clonality. The discriminatory power of PFGE depends on the number and distribution of restriction sites throughout the genome, including extra-chromosomal DNA, which define the number and sizes of bands in the profile, and can be increased by using different or combinations of restriction endonucleases. Within and between laboratory '*reproducibility and repeatability*' of results, based on the experience gained in the context of PulseNet International and PulseNet Europe , can be high, but the technique may be thought as laborious and time consuming. PFGE may require several days for completion, with time increasing with the number of restriction enzymes used. '*Harmonised standards'* are available, with the exception of a completely harmonised nomenclature, although for *Salmonella* a harmonised and agreed nomenclature is used within the EU (see Chapter 4). Nevertheless, achieving a uniform international nomenclature for '*future harmonisation*' should be possible.

3.1.4. Specific gene characterisation

Specific gene characterisation describes the selection of a panel of genetic markers, usually a set of virulence or antibiotic resistance related genes, for typing purposes. Isolates are compared using the same set of markers for their presence, and the pattern of genes is used as a typing 'profile' to differentiate isolates (Foley et al., 2009).

An example of specific gene characterisation that is commonly used is STEC virulence gene typing by PCR or microarray, in which STEC strains are compared by analysing a panel of known virulence genes for their presence or absence (see chapters 2 and 4). This approach can be used on any species, provided a known set of virulence markers are described.

The ease and speed of specific gene characterisation makes it a good candidate to combine with other existing typing methods. It can also be run at low cost.

The results of specific gene characterisation can be presented by indicating the presence or absence of a number of selected genomic targets. The targets could be either genes, repetitive elements, phage related sequences or any other stable genetic locations providing inter-species polymorphisms. The targets may be detected using multiplex-PCR (standard or real-time) or any suitable array-technology. The common result will be an array of loci with variable presence between the isolates typed. Usually, a zero '0' is used to denote absence of marker and one '1' is used to denote presence. This will give an array of ones and zeros, for example '1100011110', which can be used as a 'binary typing code'. This is why this methodology is also referred to as 'binary typing'.

Specific gene characterisation has a moderate to high '*discriminatory capability'* depending primarily on the number of targets sought. The discriminatory capability also depends upon the panel of markers selected, which needs careful consideration. The methodology includes validated technology, and thus displays high 'reproducibility and repeatability', which may be effected by the size of the assay as large arrays may be more difficult to standardise and interpret. There is not as yet '*harmonisation at international level',* apart from software tools that may be employed internationally. Nevertheless, there is '*potential for international harmonisation*' if appropriate quality control strains are used. The time frame of obtaining typing results is usually within 24 hours of receiving isolates.

Mobile genetic elements such as plasmids, transposons, insertion sequences and integrons can also be characterised. Special care should be taken since these elements may not be stable and may be lost during storage or culture in some cases, thus compromising reproducibility when assessing the clonality of the strains. Plasmid characterisation is very useful in the study of spread and transfer of antimicrobial resistance genes. Identification and molecular characterisation of plasmids encoding antibiotic resistance and virulence genes may also be helpful in outbreak situations.

3.1.5. Multiple-Locus Variable number tandem repeat Analysis (MLVA)

All bacterial MLVA-assays simultaneous measure the length of variable number of tandem repeat (VNTR) loci by PCR amplification and electrophoresis, and use this information to create a genotype to distinguish between isolates of the same species.

MLVA has several advantages: it has a high index of discriminatory power, which can be easily adjusted by inclusion or exclusion of loci to be investigated; handling of pathogenic bacteria is low, which increases laboratory safety; rapidity, as both PCR and electrophoresis times can now be greatly reduced due to improved technology. The MLVA methodology is additionally able to adapt to the new developments in separation technology (e.g. microcapillary electrophoresis linked with sequencing). Improved or equal resolution compared to PFGE is obtained when capillary electrophoresis is employed. MLVA patterns are also easier to interpret than banding patterns generated by methods such as PFGE, AFLP and REA, and especially so when multiple MLVA loci are coloured with different fluorescence dyes which identify bands that correspond to a specific locus. A drawback is the fact that MLVA-assays are very specific, thus a new assay is generally needed for each species of organism (or even variants such as serovar of the same species) under study. Another complication is the occurrence of strains presenting null allele or multiple alleles for a particular repeat locus, which may complicate or bias the type comparison.

Multi-locus variable number of tandem repeat analysis (MLVA) have been evaluated as an alternative to PFGE for STEC O157 (Noller et al., 2003; Hyytia-Trees et al., 2006), STEC O26 (Miko et al., 2010), *S.* Typhimurium (Larsson et al., 2009) and *S.* Enteritidis, and standard protocols are freely available via Pulsenet[9].

MLVA typing has a high '*discriminatory power*' for *Salmonella*, STEC and *L. monocytogenes* but not for *Campylobacter*. MLVA samples a limited portion of the genome related to areas containing tandem repeats. Only *S.* Typhimurium MLVA has so far been validated for international 'reproducibility and repeatability', and results indicate high reproducibility and repeatability when strict guidelines and a reference strain collection are used. MLVA allows direct digital storing of results as discrete-character numeric data. For inter-laboratory comparability and the correct assignation of the numeric profile, calibration of measured fragment sizes has to be performed in each laboratory (Larsson et al., 2009). A proposed standardisation scheme also exists for *S.* Enteritidis. Thus, **'***international harmonisation*' appears well advanced, in particular for *S.* Typhimurium. Furthermore, the '*potential for future international harmonisation'* for *Listeria* and STEC, but not for *Campylobacter*, should be possible based on the experience with *S.* Typhimurium. MLVA results can be obtained within 24 hours of receiving isolates.

Recently, a proof-of-concept study for successful inter-laboratory comparison of MLVA results was published involving 20 international laboratories which MLVA typed 15 strains of *S.* Typhimurium. The publication provided a comprehensive tool that enables laboratories to compare the vast majority of their MLVA results regardless of what hardware, software, primers and conditions they are using (Larsson et al., 2013). A publication on the development and application of MLVA methods in general as a tool for inter-laboratory surveillance was additionally published, and this paper proposes an international consensus on the development, validation, nomenclature and quality control for MLVA used for molecular surveillance and outbreak detection based on a review of the current state of knowledge (Nadon et al., 2013).

[9] Further details available at: http://www.pulsenetinternational.org/assets/PulseNet/uploads/mlva/2013%20updates/PNL27_MLVASalmEBeckmanProtocol.pdf (last visited on 11/12/2013)

3.1.6. Sequence-based typing methods

3.1.6.1. Single Locus Sequence Typing (SLST)

SLST describes the sequencing of a single gene or genetic locus, which displays enough polymorphism to be used in a typing scheme. Usually one single locus is sequenced and compared between strains to determine genetic distance. The SLST method thus entails the same operational steps as running Multi locus sequence typing (MLST, see below) the only difference is the number and selection of the target loci. Equipment and analysis software used will in most instances be the same. Sequencing of the flaA short variable region (SVR) may be used for typing of *Campylobacter* (Meinersmann et al, 1997). This provides a good discrimination within *C. jejuni* and *C. coli*, and an international nomenclature is established (via the pubMLST database). The flaA-SVR is often used as an additional locus to the seven MLST loci to improve the discrimination of MLST.

For STEC, Scheutz et al., (2012) compared *stx* sequences and grouped them according to genetic relatedness. Based on those results, a protocol was established for the subtyping of both *stx*1 and *stx*2 using PCR. This protocol was tested against a panel of 62 STEC reference strains especially established for this study, a small selection of 162 clinical isolates, and all 42 strains from the German HUSEC collection (Scheutz et al., 2012) (see Chapters 2 and 4). This protocol has proven highly valuable in assessing STEC pathogenicity, and a standardized nomenclature is described.

SLST has a high '*discriminatory power*' for subtyping known STEC *stx* variants, and moderate capability for *Campylobacter* spp. *flaA* SVR typing. For *Salmonella* and *Listeria*, SLST is not commonly used. '*Reproducibility and repeatability*' are considered high but '*current international harmonisation*' would require establishment of international SOPs and EQA procedures, although harmonised nomenclature and data management tools are already in place. These could be developed without major difficulties, so the method could have a high capability for '*future international harmonisation*' SLST methodology is well proven, and typing results in most cases will be available with 24 hours.

3.1.6.2. Multi locus sequence typing (MLST)

MLST indexes sequences variation at a number (usually seven) genetic loci distributed around the chromosome (Maiden, 2006). These are ideally 'housekeeping' genes, i.e. genes encoding enzymes that are involved in primary metabolism of the organism in question and which are therefore present in all isolates. Such genes are also subject to stabilising selection for conservation of metabolic function. In other words, any sequence changes occurring in these genes must be compatible with survival in the face of competition in the organism's host or environmental niche, if detrimental such changes will be lost from the population by selection against them. With this method an allelic profile or sequence type (ST) is created for each pathogen. The STs are also assigned unique arbitrary identifiers so that the sequence variation can be summarised as a single number. The existence of web-accessible databases of allele definition, ST and isolate data enables the unambiguous comparison of data collected in different laboratories. A number of analysis approaches can be used to examine structure within MLST datasets and establish relationships among STs which are crucial for identifying membership of clonal complexes.

Since its introduction in 1998, MLST has been widely deployed for typing bacteria in general and food-borne pathogens in particular and schemes exist for each of the four pathogens considered (i.e. *Campylobacter coli and C. jejuni*, *E. coli*, *Salmonella enterica*, and *Listeria monocytogenes*). The approach was devised to account for high levels of recombination observed in many bacterial populations, but is effective for organisms that exhibit the entire range of population structures from clonal through to non-clonal. A major advantage of the techniques is that in addition to reliably typing organisms, MLST provides sequence data that can be analysed in a variety of ways to study the population structure and evolution of bacterial pathogens.

MLST can additionally be targeted to virulence genes or mobile genetic elements. MLST on plasmids belonging to the incompatibility groups I, F and HI can be performed by recently developed MLST schemes (i.e. plasmid MLST databases or pMLST[10]). Such schemes permit further analysis of genetic and epidemiological relations among plasmids from different bacterial species, sources or regions. MLST assays targeting virulence genes instead of housekeeping genes are usually called Multi-Virulence-Locus Sequence Typing (MLVST) and in most cases improves the typing resolution compared to MLST and MLVST can be used to complement MLST in cases where increased resolution is needed.

Although highly successful in identifying the structure within bacterial populations, MLST suffers from a number of drawbacks. The separation of a sample from a given bacterial isolates into seven or more separate sequencing experiments requires careful management to ensure that the sequence data are accurately assembled at the end of the process. For some microorganisms, for example 'single clone' pathogens such as *S.* Typhi, seven loci do not provide sufficient resolution for useful typing, further than serovar identification, as all isolates are identical at this resolution. In other cases, such as *Campylobacter*, particular genotypes are widespread and the same MLST profile can be found in epidemiologically unrelated specimens. In both these examples additional information is required for resolving differences and similarities among isolates.

The '*discriminatory capability'* of MLST is moderate to high depending on the pathogen and gene subset typed; usually the discriminatory capability for food-borne pathogens is too low for outbreak investigations and thus additional typing data is needed when used in this context. *'Reproducibility and repeatability'* are considered to be high and '*current international harmonisation'* is well advanced, even though international SOPs could benefit from standardising an assay for each pathogen, rather than allowing different methodologies to be used.

3.2. Novel and prospective technologies

3.2.1. Whole genome mapping (WGM) analysis

WGM, or 'optical mapping' as it was named when introduced in the 1990s, produces a barcode-like 'map' of genomic restriction sites in the actual order that they appear along the genome, rather than by size distribution. The data obtained with this method are processed by dedicated computer software to generate a restriction-based map of the genome (Miller, 2013). WGM is a promising new technology for use in applications such us food-borne outbreak investigation. The resolution of the method has not been fully verified against more established methods such as PFGE, and the time to a fully finished and analysed map is about 48 hours, which is longer than many of the current technologies (e.g. MLVA, PFGE).

The '*discriminatory capability*' of WGM is believed to be high for the four pathogens of interest as it samples the whole genome, but further research is need to verify this. The same applies to its *'reproducibility and repeatability*' capacity, which is also believed to be high. *'Current international harmonisation'* is lacking, except for the availability of data management tools. The *'potential for future international harmonisation'* is unknown.

3.2.2. Whole Genome Sequence (WGS) analysis

Most of the prominent new technologies are the sequence-based. Several versions of new sequencing technologies, employing different principles, are in existence, all of which are designed with the aim of rapid sequencing of whole genomes. An often-used term is 'Next Generation Sequencing' (NGS), which is commonly used to refer to the post-Sanger and Maxam–Gilbert sequencing methods (Struelens and Brisse, 2013).

[10] Further information on pMLST available at: http://pubmlst.org/plasmid/ (last visited on 11/12/2013)

There are four approaches currently in use: (i) pyrosequencing, exemplified by the Roche 454 platform which can generate longer read lengths, but in smaller numbers and with potential miscalling of polynucleotide sequences; (ii) Illumina sequencing technology, which produces shorter sequences with very high sequence capacity; (iii) IonTorrent, which produces shorter sequences, also with a potential for miscalling polynucleotide tracts; and (iv) the PacBio SMRT sequencing system, which can produce very long sequences, but with relatively high error rates and cost. These technologies, especially those that depend on nanopores, are all in rapid development so no exhaustive review will be made here as it is likely to become outdated almost immediately. Compared to 'Sanger' sequencing all of the current methods generate individual sequence reads with high error rates: error correction is achieved with very high sequence coverage.

The end-product result is the fully sequenced genome of the pathogen. Thus, it achieves the maximum capability possible for discriminating isolates, and thus is the endpoint for DNA-based typing. Nevertheless, there are several challenges that still need to be solved mainly linked to the storage, interpretation, annotation and harmonisation of the huge amount of information provided by these technologies. In short: how to make biological sense in a day-to-day public health context out of the big data made available.

Recent technological developments have dramatically reduced the cost of determining the complete, or nearly complete, genetic information for bacterial isolates. Various approaches as mentioned above are available and these are increasingly being used applied to studies of vary large numbers of isolates (e.g. the 100K food-borne pathogen genome project[11]). Within the foreseeable future it is likely that this type of technology will be used in at many reference laboratories, although it is unclear which technologies will be routinely applied and how the data will be used.

International harmonisation on how to interpret DNA-sequence information will be paramount. International agreement on storage and sharing of data must also be addressed, as well as agreement on how to address linked data e.g. the source and geographic location of the sequenced pathogen. This can be a very difficult topic on which to reach agreement, as country-specific legislation will most likely restrict full sharing of information. There is no international agreement on the harmonisation and storing of sequence data, nor is it fully agreed on how to uniformly annotate de novo sequences. The vision paper of the European Commission on the development of data bases for molecular testing of food-borne pathogens in view of outbreak preparedness[12] reflects further on some of these important issues.

Thus, the technology itself is not the cause of concern but rather how to use all the information provided in the most effective way. The risk is that failure to properly harmonise storage and interpretation of WGS data, will lead to fragmentation of data so that, for example, similar pathogens may be regarded as different due to different approaches to sequence assembly and annotation. A further risk is that data will be stored in such a manner that it will not be generally available.

Whatever technologies become routinely used, and it is possible that a range will be deployed for different applications, the major challenge will be the exploitation of WGS for epidemiological purposes. At the current time there are two ways in which these data have been used: (i) the mapping of sequence reads to a reference and (ii) de novo assembly. These approaches can be used independently or in combination. Mapping of sequences reads to a reference sequence establishes the differences between the novel sequence and a complete high-quality reference. This has been widely used but the approach depends on the availability of a suitable reference sequence and has the limitation that only variation in sequences that are present in the reference and the tested isolate can be identified. It is also time consuming and generates lists of variable nucleotides rather than contiguous

[11] Further info on the 100K food-borne pathogen genome project available at: http://100kgenome.vetmed.ucdavis.edu/ (last visited on 11/12/2013)

[12] See vision paper from the European Commission on the development of data bases for molecular testing of food-borne pathogens in view of outbreak preparedness available at: http://ec.europa.eu/food/food/biosafety/salmonella/docs/vision-paper_en.pdf (last visited on 11/12/2013)

sequences that can be independently analysed. *De novo* assembly, on the other hand, uses one or a number of algorithms to assemble the data generated into contiguous sequences which are more easily interpreted for biological inference. Very high read depths can achieve extremely high accuracy, although systematic errors can be a problem, especially in the miscalling of polynucleotide tracts with pyrosequencing and IonTorrent data. Once assembled these sequences can in theory be compared to each other directly, including K-mer type approaches, or annotated and used to extract conventional typing information or to perform gene-by-gene comparisons across the whole genome. As yet it has not been demonstrated that WGS data provides backward compatibility with PFGE or MLVA. Limited knowledge is available in relation to the technical errors that occur during sequencing and analysis and on the effect of genetic drift in the different bacterial populations over time, which may complicate the interpretation of results.

In addition to the technical issues of generating such data cost-effectively in reference laboratories, although the development of bench-top machines may ameliorate this, a major challenge is the analysis of WGS data. While the development of wholly new analytical approaches such as SNP-based typing is likely, it is perhaps more useful, at least in the immediate future, to employ WGS for recovering epidemiological typing data such as MLST, antigen gene, or virulence and antibiotic resistance-determinants from specimens.

With the advent of high throughput sequencing technologies thousands of strains may be assayed to quantitatively identify multiple markers and combinations that have a statistical association with epidemic potential or disease severity, in a similar way to what has been done in relation to genetic risk for certain cancers in people (Roukos, 2013).

The '*discriminatory capability*' of WGS is very high as it samples the whole genome, including extra-chromosomal DNA. *'Reproducibility and repeatability*' are also high. *'Current international harmonisation*' is lacking except for the availability of data management tools and annotation guidelines – but this does not provide for fully harmonised nomenclature. The *'potential for future international harmonisation'* is unknown, but should be considered high from a technical point of view.

WGS will be increasingly important in real-time investigation of food-borne diseases, especially with newer instruments offering longer reads and reduced error rates. WGS may not yet be ready to be applied routinely as a sub-typing method, but reference laboratories must participate in the future developments and assess how to most efficiently analyse and interpret WGS data in the context of molecular epidemiology of food-borne pathogens and their public health risk.

3.3. Concluding remarks on the review of the molecular typing methods

This chapter has focused on the most widely internationally used methods for molecular typing of major food-borne pathogens, and has included WGS, which seems to be the future of bacterial pathogen typing. A summary of the evaluation of the molecular typing methods presented in this chapter can be found in Appendix A.

Currently, the majority of molecular typing of food-borne pathogens is done employing one or more methods from a wide array of available molecular non-WGS methods. Non-WGS methods detailed above have been valuable in numerous applications. The relative simplicity in the analysis of the data obtained, coupled with their availability, currently make them attractive choices when simple genotyping to compare organisms is needed, and the methods will probably be used for several years to come, despite their limited discriminatory capacity in many cases. The methods have proven track records for both use and application, and for some (e.g. PFGE) extensive databases of valuable typing data have been collected. The availability and application in the EU Member States of molecular typing methods in a food, animal and feedingstuff context has been the subject of a survey carried out by EFSA in 2008 (EFSA, 2009). In order to provide an overview of the availability and application of

molecular typing methods for food-borne pathogens including the public health perspective, it is essential that the human health sector is included in future similar surveys.

Accurate discrimination of pathogenic and potential pathogenic food-borne bacteria is, and will be, an important priority to ensure food safety. This is increasingly relevant as international trade in food animals and foodstuffs, as well as travel is now widespread. Thus, the need for international harmonisation of methodology, and agreement on principles for data sharing and collaboration is paramount. It is within the field of sequencing technology where the most extensive research and development projects are being carried out. Prediction as to when WGS will become the method of choice for the majority of typing laboratories is difficult; this will depend on the development of easy to use dedicated bioinformatics tools as well as the presence of international sequence repositories. The rapid introduction of new sequencing technologies and with next generation technologies on the horizon e.g. polymerase conductance measurement sequencing (Chen et al., 2013) or sequencing by nanotechnology (Haque et al., 2013) will make it difficult to select a harmonised platform. This means that EQA procedures to ensure comparability of results will be extremely important.

In summary, the following are concluding remarks on the review of current and prospective molecular identification and typing methods:

- Molecular typing methods should ideally provide appropriate discriminatory power, reproducibility, capability for international harmonisation and reduced handling of and exposure to pathogens in the laboratories. No current typing method, whether phenotypic or molecular, complies with all these expectations.

- Several methods are often used in combination in order to obtain the resolution needed. The methods applied depend on the pathogen and on the application sought. These methods have proven track records of use and for some of them (e.g. MLST, PFGE) extensive databases of valuable typing data have been collected.

- Methods based on WGS can replace and are increasingly replacing the numerous different methodologies currently in use in human and veterinary reference laboratories, and the same methods can be used for all organisms. An essential precondition is the availability of quality control methods, to ensure the reliability and consistency of molecular data generated, coupled with high quality bioinformatics support for the analysis of the data generated.

- To properly evaluate typing methodologies, data from strain characterisation should be linked with epidemiological metadata and the strain selection must be unbiased and statistically representative of the population to be assessed.

- International harmonisation of molecular characterisation outputs by means of standardisation or appropriate quality control procedures is essential. This includes controlling the accuracy of production of DNA sequences from WGS and the further interpretations of annotation pipelines.

- Regarding WGS, limited knowledge is available in relation to the technical errors that occur during sequencing and analysis and on the effect of genetic drift in the different bacterial populations over time, which may complicate the interpretation of results.

- Modern molecular typing methods provide many opportunities for rapid and accurate determination of the genealogical relationships among bacterial isolates. Interpretation of the results generated by these methods for different public health applications requires this information to be placed in the context of the diversity, degree of genetic change (e.g. during storage of isolates or mutation during an outbreak and in reservoirs) and population structure of the particular pathogen in question. Therefore, large scale carefully co-ordinated studies are required to fully elucidate this.

- Development and improvement of international initiatives with regard to harmonised platforms for sharing of data such as those promoted by Pulsenet and ECDC/EFSA should be urgently prioritised, including the integration of WGS into these platforms.

- The development of more informative and easier to use bioinformatic tools for WGS data is needed. Otherwise the spread and adoption of WGS as an international typing tool may be restricted.

- An updated EU-wide review, similar to that carried out by EFSA in 2008, on the availability and application of molecular typing methods for food-borne pathogens including those from the public health sector is recommended.

4. The use of molecular typing methods for the detection and investigation of food-borne outbreaks of disease

4.1. Introduction

Methods for discriminating bacterial isolates are essential for the detection and subsequent investigation of putative food-borne outbreaks as well as for the implementation of measures for infection prevention and control. Typing techniques, whether phenotypic or molecular, should be able to type all the isolates studied (high typeability) and should be able to discriminate between isolates at an appropriate level (discriminatory power). The choice of molecular typing method (or methods), is therefore dependent on the problem to be solved, and the epidemiological context in which the method is going to be used, as well as the time and geographical scale of its use. For the investigation of food-borne outbreaks of disease, molecular methods *per se* can and should complement, extend, and in due course substitute for the more traditional phenotypic methods.

Requirements for molecular typing methods used in the investigation of outbreaks are also applicable to outbreak detection, as in both instances results have to be highly reproducible, validated, and easy to share between laboratories. For outbreak detection, methods should also be easy to perform and interpret, rapid and inexpensive (e.g. high-throughput) in order to allow application for routine/continuous surveillance of clinical isolates as well as isolates from food, veterinary or environmental sources. For further outbreak investigation more detailed analyses with other/complex methods can be utilised as is considered necessary (van Belkum et al., 2007; Sabat et al., 2013).

For both the detection and investigation of outbreaks, for determination of the source(s) of infection and for surveillance of the pathogens involved, the use of defined and agreed molecular typing methods for human, veterinary and food isolates is an essential prerequisite. Such methods should be portable, with agreed nomenclature for types and subtypes, common databases and agreed procedures for quality control and access to standard type strains.

4.2. Detection of food-borne outbreaks

For the purposes of this chapter, an outbreak is regarded as the existence of isolates which are indistinguishable by molecular typing or by phenotypic methods, and confirmed as related by epidemiological investigations from at least two individuals. When molecular typing is used by public health laboratories the methods must yield results with adequate stability over time to allow meaningful interpretations to be made and for the implementation of efficient infection control measures. There should also be procedures in place, by using quantifiable internal and external controls, to check and validate that the typing data are of high quality, are readily applicable to the human, veterinary and food sectors, and are reproducible and stable over time.

In relation to food-borne outbreaks, molecular typing should be able to:

- confirm the existence of like strains, including situations where epidemiological investigations have identified increases in incidence above the expected norm (i.e., contribute to exceedence reporting compared to the baseline set by routine/continuous surveillance);
- recognise the emergence of ‘new’ strains, or of old strains with new properties contributing to its pathogenicity and /or transmission;
- contribute to the identification of possible increases in morbidity and mortality in vulnerable population groups.

There are numerous examples of how molecular typing has assisted in strain identification at both national and international levels, leading to outbreak recognition and to the timely implementation of control measures (for reviews, see Fisher and Threlfall, 2005; Swaminathan et al, 2001 and 2006). There is little doubt that since the early 1980s, the application of molecular typing methods has led to

more outbreaks being identified and especially the correlation of sporadic cases spread in geography and time to a common source

4.3. Investigation of food-borne outbreaks

In food-borne outbreak investigation, an effective molecular typing method must have the discriminatory power needed to distinguish all epidemiologically-related isolates from a background of the total population of the specific species. Ideally, such a method should have the ability to discriminate closely-related isolates from human, animal and food sources. Thus molecular typing methods should be able to reveal both spread from food-production animals to derived food, and subsequent transmission to humans both of which are important in developing strategies to prevent further spread. These properties are particularly relevant to investigations which are totally reliant on molecular typing for strain discrimination. Furthermore, when strains exhibit resistance to antimicrobial drugs, which is becoming an increasingly common feature of pathogens associated with food-borne infections at the European level, then there should be agreed procedures for the identification of resistance genes, and their mechanisms of transmission.

4.4. Data needs

A molecular typing method that is used in international networks and for comparative purposes in the identification of outbreaks should produce data that are portable (i.e. easily transferrable between different systems) and that can be easily accessed *via* an open source web-based database, or a client-server database connected *via* the Internet. Additionally, a molecular typing method used for surveillance should rely on an internationally standardised nomenclature, and should preferably be applicable for a broad range of bacterial species. A clear advantage for a pan-generic typing approach is the availability of software that: (i) enables automated quality control of raw typing data, (ii) allows pattern/type assignment, (iii) implements an algorithm for clustering of isolates based on the obtained data, (iv) provides assistance in the detection of outbreaks of infections, and (v) facilitates data management and storage. To date, many different methods, both phenotypic and molecular, have been developed for the epidemiological characterisation of bacterial isolates, none of which is optimal for all forms of investigation. Thus, a thorough understanding of the advantages and limitations of the available typing methods is of crucial importance for selecting the appropriate approaches to unambiguously define outbreak strains.

For efficient source tracing in food-borne outbreak situations, the methods used for molecular typing of pathogens in food and veterinary laboratories should be comparable to the molecular methods used for similar purposes in public health laboratories. For hypothesis generation that may permit identifying the source of the outbreak good longitudinal data on baseline type frequency distribution that are sufficiently representative of the population under monitoring, whether that be food-producing animals or foods. An important method validation step is the evaluation of consistent grouping of the molecular characteristics of related isolates that form part of a common source outbreak (van Belkum et al., 2007).

Investigation of local or national outbreaks undertaken by a single reference laboratory may not challenge the methodology to the same extent as outbreaks involving multiple countries. In the latter, exchange of harmonized molecular data between laboratories coupled with comparison of such results over time by type allocation based on standardized nomenclature is essential. Ideally, such results should be included in cumulative, shared and curated databases across relevant sectors (human, veterinary and food). A precondition for such a system is common agreement both among sectors and internationally on the harmonization and standardization of methods and the willingness to share results, including related epidemiological information, at appropriate levels. A further essential precondition is the availability of agreed quality control methods, to ensure the reliability and consistency of molecular data generated. In some circumstances a two-step approach could be considered, e.g. a fast, high-throughput molecular method for the daily surveillance in public health as well as food/animals and a more resource demanding typing method for highly discriminatory typing of possible outbreak-related isolates, selected on the basis of the first method.

4.5. Past and current experiences employing molecular typing methods

The PulseNet network for the molecular subtyping of food-borne bacterial pathogens, introduced in the USA in the early 1990s for the investigation of outbreaks of STEC O157 (Swaminathan et al., 2001) and into Europe shortly afterwards (Peters et al., 2003; Gerner-Smidt and Scheutz, 2006) is undoubtedly the most successful international molecular typing network to date. The method is based on fragmentation of chromosomal DNA using rare-cutting restriction endonucleases, followed by the resolution of resultant DNA fragments by pulsed field gel electrophoresis (PFGE) (see Chapter 3). The method has been extended to encompass food-borne pathogens such as *Salmonella* (Swaminathan et al., 2001; Liebana et al., 2002; Peters et al., 2003), *Listeria monocytogenes* (Graves and Swaminathan, 2001; Graves et al., 2005; Martin et al., 2006; Felix et al., 2012)) and to a lesser extent *Campylobacter* (Ribot et al., 2001; Sahin et al., 2012). PFGE for food-borne pathogens has been standardised both nationally in the USA, and internationally through various PulseNet International networks in most continents (Swaminathan et al., 2006) to allow for inter-laboratory comparison. The standardised PFGE protocols are now in use in laboratories worldwide for surveillance of food-borne infections, and for outbreak detection and source tracing of outbreaks of both food-borne and nosocomial bacterial pathogens.

In Europe, the standardisation of PFGE for *Salmonella* was initially undertaken within the Enter-net [13]network (Peters et al., 2003), under the auspices of the Salm-gene project, resulting in the rapid EU-wide identification of many outbreaks between 2002 and 2008 (Fisher and Threlfall, 2005), and more recently by PulseNet Europe[14] (Pezzoli et al, 2007), following the consolidation of the SalmGene molecular databases with databases of profile types from food animal isolates (Denny et al., 2007). The *Salmonella, Listeria* and STEC PFGE profiles from the PulseNet Europe database are now incorporated into the pilot Molecular Surveillance System (MSS) established by ECDC in 2012 as part of TESSy. From 2012 *S.* Typhimurium MLVA data is also part of the MSS.

Exchange of molecular typing data has been instigated between some Scandinavian countries for ad-hoc investigations (Bruun et al., 2009). Within the UK a network for sharing *Salmonella enterica* molecular typing data between public health and veterinary agencies has been in existence for at least two decades. (Liebana et al, 2002, 2004; Lawson et al., 2004). At the international level, in 2010 a collaborative network involving 15 human and veterinary laboratories in 10 countries unequivocally demonstrated the European-wide dissemination of closely-related strains of multidrug-resistant monophasic *S*. Typhimurium in food animals and humans (Hopkins et al, 2010). For *Campylobacter*, an international database of MLST types has been established (Colles and Maiden, 2012) and is widely used for outbreak investigations and source attribution studies (see 4.6.3 below).

In the food and veterinary sector a questionnaire survey on the availability of molecular typing methods for the main food-borne pathogens in animals, food and feedingstuffs in EU MSs was performed in 2008 (EFSA, 2009). The conclusions were that molecular typing was performed at least on some occasions for all the pathogens included in the investigation (*Salmonella* spp., *Campylobacter* spp. (thermophilic), STEC, *L. monocytogenes* and *Staphylococcus aureus*), but only few countries typed isolates on a routine basis. PFGE was used by most countries for the molecular typing of all pathogens listed above, whereas fewer countries used other methods, such as MLVA, MLST, Spa typing, ribotyping, RFLP, AFLP, *fla* typing and SSC*mec* for *S. aureus*. Information on storage and share of data among laboratories, countries and/or sectors was not collected in the survey. Furthermore, the development of an EU-wide molecular typing database for food and veterinary isolates is underway following a mandate from the European Commission to EFSA and ECDC[15].

[13] Further information on Enter-net available at: http://ecdc.europa.eu/en/activities/surveillance/pages/enter-net.aspx

[14] Further information on PulseNet Europe available at: http://www.pulsenetinternational.org/networks/europe

[15] For further information on the mandate from the European Commission to EFSA and ECDC on the building of a molecular typing database visit: http://registerofquestions.efsa.europa.eu/roqFrontend/questionLoader?question=EFSA-Q-2013-00250 (last visited on 11/12/2013)

If submission of molecular typing data to a central database (as in the PulseNet USA and PulseNet Europe databases, and the EU TESSy-MSS) is done in real-time by laboratories covering different regions or countries, geographically-dispersed outbreaks of disease can be detected. As mentioned earlier, this requires the generation of fully comparable typing data from laboratories typing isolates from cases of human infections. Similarly, for outbreak investigations and source tracing comparable methods must be used in the relevant sectors, i.e. laboratories typing isolates from cases of human infection, food animals and food.

4.6. Current molecular methods used for outbreak detection and outbreak investigation

When choosing methods for the molecular typing of food-borne pathogens, for the identification of genes encoding antibiotic resistance and/or virulence, and the characterisation of the elements responsible for the dissemination of such genes, both the short-and long-term perspective should be considered. For the immediate future, the present experience and capacity in European laboratories should be taken into account

4.6.1. *Salmonella* spp.

For serotyping of *Salmonella*, a molecular approach based on detection of the specific detection of genes coding for the relevant O- and H-antigens has been developed (Fitzgerald et al., 2007; McQuiston et al., 2011) but this approach has replaced conventional serotyping only to a limited degree. In addition, Achtman et al. (2012) have recently proposed an MLST-based, 'e-burst' method to replace conventional serotyping, but further validation is required before this method can be adopted internationally. For subdivision within *Salmonella* serovars, the relevant molecular methods are PFGE (see above), MLVA for *S.* Typhimurium (Lindstedt et al., 2004; Best et al., 2007, 2009; Petersen et al., 2011; Prendergast et al., 2011a; Paranthaman et al., 2013);and *S.* Enteritidis (Boxrud et al., 2007; Cho et al., 2010; Hopkins et al., 2011; Mossong et al., 2012; Dewaele et al., 2012; Lindstedt et al., 2013), clustered regularly interspaced short palindromic repeats (CRISPR) for a range of *Salmonella* serovars (Liu et al., 2011; Fabre et al., 2012) including *S.* Enteritidis (Liu et al., 2011b; Shariat et al., 2013a), *S.* Typhimurium (Dimarzio et al., 2013) and *S.* Newport (Shariat et al., 2013b). A major drawback to CRISPR typing is that the method is used mainly in France but not internationally.

MLVA specific for *S.* Typhimurium and monophasic variants (Lindstedt et al), 2004 has now been implemented as a standard molecular typing method for the pilot phase of the MSS implemented by ECDC in 2012 as part of TESSy.

4.6.2. *Listeria*

For *Listeria*, a PCR-based method for the serogrouping of *L. monocytogenes* is validated and has a defined nomenclature that corresponds to the conventional serotyping scheme (Doumith et al., 2004). This gel-based PCR is well established in reference laboratories and included in the two major European EQA programmes organised by the EU-RL for *L. monocytogenes* (food laboratories) and the ECDC FWD network (public health laboratories), respectively. Recently, a comparable real-time PCR method targeting the same genes has been propounded by Vitullo et al. (2013), but further validation is required before this can be adopted at the level of the reference laboratory. For subtyping, the most commonly used molecular method in the EU is presently PFGE (see 4.5 above), and an EU-wide study of PFGE-generated typing results for *Listeria* is underway. MLVA has also been successfully evaluated (Lindstedt et al., 2008; Chenal-Francisque et al., 2013) and AFLP (Ripabelli et al, 2000) is the method of choice in the UK. Automated ribotyping has also been used for this purpose in Austria (Grif et al., 2006). A recently developed multi-virulence-locus sequence typing (MVLST) method has shown improved discriminatory power for subtyping genetically diverse *L. monocytogenes* isolates and identification of epidemic clone isolates associated with two recent USA multistate listeriosis outbreaks, and accurately identified three previously known epidemic clones and detected another epidemic clone in serotype 4b of *L. monocytogenes* (Chen et al, 2007).

4.6.3. STEC

For STEC consistent nomenclature and subtyping strategies especially in relation to the identification of virulence genes are of primary importance for surveillance and outbreak detection/investigation as well as for predicting the risks associated with particular STEC infections. Present methods for the identification of virulence genes in STEC, and for the classification into seropathotypes (EFSA BIOHAZ Panel, 2013) are based on the identification of virulence genes (typically by PCR assays) and by O-serogrouping. In this respect a PCR protocol for the subtyping of the Stx-encoding genes has been developed and evaluated by means of a multicentre study (Scheutz et al., 2012), and also locally in the UK (Jenkins et al, 2012). For the molecular subtyping of STEC, a standardised PFGE protocol (Ribot et al., 2006) has been widely distributed through the PulseNet networks (see above), and feasibility studies have been undertaken in both the USA and Europe (Gerner-Smidt and Scheutz, 2006; Ribot et al., 2006; Gerner-Smidt et al., 2005).

More recently SNP typing has been used in the USA to investigate a recent cluster of *E. coli* O157:H7 infections attributed to salad bar exposures and romaine lettuce (Turabelidze et al., 2013). In this putative outbreak a subset of cases denied exposure to either source, although PFGE and MLVA suggested that all isolates had the same recent progenitor. Interrogation of a preselected set of 3 442 673 nucleotides in backbone open reading frames (ORFs) identified only 1 or 2 single nucleotide differences in three of 12 isolates from the cases who denied exposure. The backbone DNAs of 9 of 9 and 3 of 3 cases who reported or were unsure about exposure, respectively, were isogenic. These SNP typing-based investigations confirmed the involvement of the subset of cases who denied exposure to the vehicles of infection, and demonstrated that backbone ORF SNP set sequencing can offer pathogen differentiation capabilities above and beyond that provided by PFGE and MLVA in outbreak investigations.

The major outbreak of STEC O104:H4 in 2011 (Frank et al., 2011a, b; Karch et al., 2012) resulted in recognition of a new paradigm for investigating STEC outbreaks. In particular, WGS suggested that the clinical characteristics of the outbreak strain were due to the unique combination of virulence factors carried by the pathogen (Rasko et al., 2011; Frank et al, 2011, b).

4.6.4. *Campylobacter*

Because of the high incidence of evolutionary diversity, genome plasticity and rapidity of change within the *Campylobacter* genus (see Chapter 2), molecular typing of *Campylobacter* for continuous surveillance and outbreak detection is rarely undertaken. Nevertheless for outbreak investigations a number of different methods have been used.

PFGE has been standardised for *Campylobacter* spp.(see 4.5 above), and has been used in localised outbreak investigations (Fitzgerald et al, 2001).

MLST (Dingle et al., 2001, 2005) has become the most commonly used molecular method for the subtyping of *Campylobacter*. An international database of MLST types has been established (Colles and Maiden, 2012) and is widely used. MLST was proven useful in an outbreak of campylobacteriosis in the UK in 2011 associated with duck livers (Abid et al., 2013), and in the attribution of campylobacter infections in northeast Scotland to specific sources (Strachan et al., 2009). *flaA* RFLP and *flaA*-SVR sequencing have also proved useful in some outbreak investigations (Wassenaar et al., 2009; Magnusson et al., 2011)

More recently whole genome MLST has been used to provide definitive characterization of *C. jejuni* and *C. coli* isolates in real time from cases of human infection in the Oxford region of the UK over a four-month period in 2012 (Cody et al., 2013). In this study typing and phylogenetic information was extracted and comparative analyses were performed for various subsets of loci, up to the level of the whole genome. The study demonstrated that clinically and epidemiologically informative data can be extracted from whole-genome sequence data in real time with straightforward, publicly available tools, and provide improved resolution (i) among *Campylobacter* clonal complexes and (ii) between very

closely related isolates. Additionally, these analyses confirmed that most isolates were unrelated, although some single-strain clusters were identified.

4.6.5. Antibiotic resistance genes, virulence, and associated elements

The previous sections focus has been on the detection and tracing of the specific aetiological agent(s). An alternative approach for outbreak detection and investigation is to focus on the analyses that identify specific antibiotic resistance and/or virulence genes or on those genetic elements carrying these genes. This approach, which cuts across bacterial genera, is further discussed in Chapter 6 – 'The use of molecular typing methods in the early identification of food-borne organisms with future epidemic potential and their integration in risk assessment'.

4.7. Potential use of new technologies

The long term perspective should undoubtedly focus on development of typing methods based on data generated by WGS. In this respect WGS should be regarded as the technique that produces data for further interpretation (Allard 2013). The actual ways of interpreting such data needs to be developed and validated for the purpose of 'typing' for public health purposes including surveillance, outbreak detection and investigation.

To date, WGS has been used for research and for retrospective investigations of localised outbreaks of STEC (e.g. STEC O157 (Underwood et al., 2013), STEC O104:H4 (Mellmann et al, 2011; Frank et al., 2011b; Karch et al., 2012)) and *Salmonella* (e.g. *S.* Newport (Zhang et al., 2013; Cao et al., 2013), *S.* Typhimurium DT 104 (Mather et al, 2013)), for the identification of resistance genes in specified populations (SSI and DTU, 2013; Zankari et al., 2012, 2013) and for comparisons between *Campylobacter* sequence types (Biggs et al., 2011). Recently, a comparative WGS analysis of *S.* Enteritidis strains causing egg-related outbreaks led to the discovery of new genetic targets useful for distinguishing isolates otherwise indistinguishable by current molecular markers and linking genomic variants to specific egg production farms (Allard et al., 2013).

A variety of different bioinformatics-based methods of performing data analysis are already described (for review see Cheung and Kwan, 2012). Data analysis can produce outputs similar to present typing methods, e.g. MLST (possible extended to additional loci), identification of antimicrobial resistance genes, and the presence of /variation in virulence genes. Calculation of phylogenetic trees, e.g. based on SNPs in the core genome, is a less 'definitive' method in the sense that the tree and the relationship between specific strains may change when additional isolates are included in the analysis. Resultant data may be useful for the detailed analysis of possible outbreak isolates and for the determination of evolutionary relationships, but may be complicated for long term surveillance. There is now an urgent need for coordinating the development and use of WGS, and validation of WGS-generated data, coupled with agreement on how to use and interpret such data for outbreak detection and surveillance. In particular, agreed systems of nomenclature of 'types' are important before such methods can be used internationally for outbreak investigations.

4.8. Concluding remarks

The essential requirements for harmonisation, standardisation and common agreements have a major drawback in that change to new methods may be resource-demanding and time-consuming. By nature, sequence-data are platform-independent and fully portable, and therefore do not require instrument or protocol standardization to the same extent as DNA fingerprinting methods. Nevertheless, since different platforms and different sequencing depths/lengths may have an impact on the data quality, harmonisation, standardisation and common agreements have to be obtained in relation to methods for analysis and interpretation of results. The infrastructure for molecular data from cases of human infection established in PulseNet and PulseNet International, and taken forward in Europe through Enter-net, PulseNet Europe and more recently, TESSy-MSS (see above), can be used for sharing of results obtained by other methods and MLVA is already relevant in this context. Similar systems for molecular data for organisms from animals and foods have been developed and implemented under the

auspices of the EU Commission-funded Med-Vet-Net Network of Excellence (Hopkins et al, 2011 – see above). A current focus is on how molecular technology systems can be adjusted for implementation of WGS-based methods, since there is a danger that multiple systems with imperfect comparability will emerge in different organisations as a result of local preferences and commercial pressures.

There is now a requirement for knowledge building and consensus on parameters such as, read lengths, the number SNPs which can be regarded as sufficient to define isolates as epidemiologically-unrelated, on the platforms to be used for data generation, on the storage and curation of WGS data, on quality assurance, and on accessibility and interpretation of data. There will undoubtedly be differences in the pace of change to WGS-based methods in different laboratories and countries, ,which will challenge the possibility of aligning results from WGS with current molecular methods as well as the need for sustain historical knowledge based on results from traditional techniques. The future challenge brought about by WGS technology will be to maintain comparable, usable and validated molecular typing data from all countries and sectors in the coming years, and to interconnect with relevant epidemiological databases.

In summary, the following are concluding remarks on the use of molecular typing methods for food-borne outbreak detection and investigation:

- Detection of outbreaks and their investigation in real-time would be enhanced by the generation of fully comparable molecular typing data from human, veterinary and food laboratories prior to submission to a central or connected databases.

- Many different methods have been developed for the molecular characterisation of bacterial isolates of *Salmonella*, STEC, *Campylobacter* and *Listeria monocytogenes* associated with food-borne outbreaks of disease, none of which meet all demands or is optimal for all forms of investigation.

- Some molecular typing methods (e.g. MLST, PFGE, MLVA) have been harmonised to a greater or lesser extent for the purpose of outbreak detection and investigation. The international development of harmonised platforms for WGS-generated data should be encouraged.

- Cross-sector and international coordination of the method validation is required as a priority.

5. The use of molecular typing methods for food-borne source attribution

5.1. Requirements and role of typing for food-borne hazards source attribution

Efforts to quantify the importance of specific sources (including foods) and animal reservoirs for human illness have been gathered under the term ‘source attribution’ or ‘human illness attribution’ (EFSA, 2008; Pires et al., 2009). Source attribution has been defined as the partitioning of the human disease burden of one or more food-borne infections to specific sources, where the term source includes animal reservoirs and vehicles, e.g., foods (Pires et al., 2009). Knowledge of the most important sources and animal reservoirs will ensure a more targeted control of the disease in question and support risk managers in their decision of allocating resources to achieve the highest possible public-health benefit. Source attribution is therefore regarded as an important tool in the process of identifying and prioritizing effective food safety interventions (Havelaar et al., 2007).

Methods for source attribution of food-borne diseases include microbiological approaches, epidemiological approaches, intervention studies and expert elicitations. The applicability of each method to address a given question will depend on a variety of factors, such as data requirements and availability, pathogen characteristics, and the type of intervention aimed for. Each method presents strengths and limitations, and the utility of each will depend on the public health questions being

addressed (Batz et al., 2005). In this chapter, the focus will be on the role and use of microbial subtyping for source attribution of food-borne hazards.

5.1.1. The microbial subtyping approach

The microbial subtyping approach involves characterization of isolates of the pathogen by phenotypic and/or genotypic subtyping methods. The principle is to compare the distribution of subtypes in potential sources (e.g. animals and food) with the subtype distribution in humans. The microbial subtyping approach is enabled by the identification of strong associations between some of the dominant subtypes and specific reservoirs, providing a heterogeneous distribution of subtypes among the sources. The approach also requires a collection of temporally and spatially related isolates from various sources and humans, followed by the application of discriminatory subtyping methods (Pires et al., 2009).

Subtyping of food-borne bacteria has for the most part relied on biochemical markers (biotyping), immunological reactions (serotyping) or bacteriophage susceptibility (phage typing). Such phenotypic methods have proven useful to make inferences about the main reservoirs for human infections for particular food-borne pathogens (e.g. *S.* Enteritidis in poultry, *Y. enterocolitica* O:3 in pigs and *E. coli* O157 in ruminants). These methods are still valid, but are increasingly being replaced by or supplemented with molecular methods based on characterisation of the bacterial DNA.

The most commonly applied DNA-based methods either generate banding patterns (e.g. PFGE, AFLP, MLVA) or DNA sequences (e.g. MLST). Often these methods have been developed specifically to characterize very closely related isolates (e.g. in outbreak investigations) or, in contrast, to compare very distant related isolates (e.g. in evolutionary studies) (Barco et al., 2013). The former methods normally investigate fast-evolving genetic markers, whereas the latter methods target the conserved and slowly evolving core genes. In source attribution studies, the appropriate level of discrimination for most pathogens will lie somewhere in between these two applications and will among others depend on the clonality of the pathogen investigated. For highly clonal pathogens like e.g. *S.* Typhimurium DT104, the subtyping method needs to be fairly high in discriminatory ability (Son et al., 2013), whereas for *Campylobacte*r spp. less discriminatory methods are needed.

The value of molecular subtyping methods for source attribution has not been fully examined, and will undoubtedly challenge the most wanted and optimal strategy: 'one typing methods that fits all needs'. This is because highly discriminatory methods are not necessarily the best solution for source attribution, where we are not looking for a single source for a particular cluster of human cases, but rather want to relate groups of bacterial strains with particular reservoirs/sources and then attribute human sporadic cases to these sources. Since it is not possible to design a practical and economically feasible surveillance system from where we can expect to identify direct links between sporadic human cases and the causative sources, we need a process that allows for some genetic diversity between strains from human and food sources, but only to the degree so that it can still be assumed that they are epidemiologically related. In other words, the methods chosen should be discriminatory enough to identify a link between human isolates and their sources, but they should not be too discriminatory, so that a true epidemiological association between isolates might be missed (Barco et al., 2013).

Subtyping methods for source attribution are typically applied on bacterial isolates collected through surveillance programmes from a variety of sources, and the typing is performed in many different laboratories. Reproducibility and standardisation are therefore also essential criteria for the chosen methods in order for the results to be comparable between laboratories, and consequently useful for source attribution analysis. Finally, it is important that the methods have a high degree of typeability, meaning that a high percentage of the typed isolates can be assigned a distinct and definitive subtype (Hunter and Gaston, 1988).

The microbial subtyping approach is limited to pathogens that are heterogeneously distributed among the reservoirs. This makes it appropriate for pathogens, which are clonally distributed and presents at least some host-associated subtypes. Because of this human cases are attributed to the reservoir level, whereas the actual pathway through which the pathogen finally reaches a human host is not elucidated. As an example: cattle are the reservoirs for *S.* Dublin, but the relative importance of different pathways within this reservoir (e.g. milk, veal, beef or direct contact) cannot be estimated using microbial subtyping alone. The microbial subtyping approach is quite data intensive, requiring a collection of isolates from all (major) sources that should to the extent possible represent what the human population is exposed to i.e. the isolates from humans and sources should be related in time and space. This is a fundamental requirement, but it often violated due to the lack of systematically collected surveillance data and/or the subsequent application of standardised subtyping methods. Because of this requirement, the approach is facilitated by an integrated food-borne disease surveillance programme focused on the collection of isolates from the major food animal reservoirs and their products and from humans (Pires et al., 2009; Smid et al., 2013).

An important advantage of the microbial subtyping approach is that it allows for the identification of the most important reservoirs of the zoonotic agent, assisting risk managers to prioritize interventions and focus control strategies at the animal level. Finally, when applied on a regular basis, the microbial subtyping approach allows for the analysis of the dynamic spread of organisms and trends in the most important sources of disease over time.

5.1.2. Available source attribution models

The most commonly used source attribution models using subtyping can be divided into two classes: 1) the frequency-matching models and 2) the population genetic models.

5.1.2.1. Frequency-matching models

The principle of comparing the distribution of subtypes found in animal and food sources with those found in humans to make inferences about the most important sources of human disease has been applied for decades by several research groups (e.g. Rosef et al., 1985; Sarwari et al., 2001). In the Netherlands, Van Pelt et al. (1999) described a simple deterministic model for estimating the number of human reported *Salmonella* cases per subtype (serotyping and phage type) and source by multiplying the number of reported cases per subtype and the relative occurrence of the same subtype in the different sources. This approach assumes that all *Salmonella* subtypes and sources have an equal probability of causing human disease. However, it is widely recognized that *Salmonella* subtypes differ in their ability to cause disease in humans, also often leading to different levels of severity (EFSA Panel on Biological Hazards, 2012). From surveillance data, it can also be seen that it is difficult to identify a linear relationship between the occurrence (prevalence) of a particular *Salmonella* subtype in food and the occurrence of reported human cases (Hald et al., 2004). Although these two latter points have mainly been described for *Salmonella*, they are also believed to be true for other food-borne pathogens. Models based on frequency matching should therefore be able to account for variations between pathogen subtypes and sources. However, this is not straightforward, since these variations are seldom quantifiable.

Using data from the integrated Danish *Salmonella* surveillance program, a stochastic Bayesian model was developed to quantify the contribution of each of the major food animal sources to human *Salmonella* infections (Hald et al., 2004; Hald et al., 2007; Pires and Hald, 2010). This model is based on phenotyping (i.e. serotypes, phage types, and antimicrobial resistance profiles) and attributes domestically acquired laboratory-confirmed human *Salmonella* infections caused by different *Salmonella* subtypes as a function of the prevalence of these subtypes in animal and food sources and the amount of each food source available for consumption. The principle behind the model is that subtypes exclusively or almost exclusively isolated from one source are regarded as indicators for the human health impact of that particular source, assuming that all human infections with these subtypes originate only from that source. Human infections caused by subtypes found in several reservoirs are then distributed relative to the prevalence of the indicator types. So, like the Dutch model, the Danish

model is based on comparing frequency distributions, but the model also compares impact across subtypes and because of the Bayesian approach, the model is able to estimate the relative impact of *Salmonella* subtypes and the included food sources. One limitation is that these subtype-dependent and source-dependent parameters estimated to account for these relative impacts are arbitrary and can best be described as multiplication factors. This helps the model to arrive at the most likely solution given the observed data (Hald et al., 2004; David et al., 2012). The specific parameters are consequently difficult to interpret and consistency between models using different datasets is not always seen. Another limitation of the Danish model is that it does not include uncertainty around the prevalence estimates, meaning that it relies on data from an intensive and integrated surveillance system providing accurate and precise subtype-specific prevalence estimates.

The Danish model has been adapted to attribute human salmonellosis in other individual EU countries (Pires et al., 2008; Wahlström et al., 2010; Valkenburgh et al., 2007; David et al., 2013), in EU as a whole (Pires et al., 2011; Hald et al., 2012), in the United States (Guo et al., 2011), New Zealand (Mullner et al., 2009), and Japan (Toyofuku et al., 2011), as well as for attribution of other food-borne pathogens e.g. *L. monocytogenes* (Little et al., 2010) and *Campylobacter* (Mullner et al., 2009; Boysen et al., 2013).

Some of the more comprehensive adaptations to the Danish model were developed by Mullner et al. (2009) and included estimation of uncertainty around prevalence parameters, which make the model more suitable in situation with less intensive surveillance data. Modification of the prior specification for the subtype-dependent parameters was also done in order to avoid assuming similarity between some subtype parameters, which was done in the original Danish model to circumvent over-parameterisation. Similarly, David et al. (2013) have proposed to specify the subtype-dependent factors for subtypes occurring in only a single source as a constant value thereby improving convergence of the model. Finally, Mullner et al. (2009) and Pires and Hald (2010) included time period as a dimension in the model, thereby making the estimation of the subtypes-depending parameters more robust.

Frequency matched models can employ both phenotypic and genotypic data. In fact, subtypes can be defined through any combination of phenotypic and/or genotypic data. Until now, frequency-based attribution models for *Salmonella* have only used phenotypic data, whereas molecular methods have also been applied for *Campylobacter* (MLST) (Mullner et al., 2009; Boysen et al., 2013). For attribution of *Listeria monocytogenes* a combination of phenotyping (serotyping) and molecular typing (AFLP) was used (Little et al., 2010). For *Salmonella*, the usefulness of molecular methods (e.g. MLVA and WGS-based methods), perhaps in combination with phenotypic methods, for source attribution still needs to be explored.

5.1.2.2. Population genetic models

Driven by the recent development of molecular typing techniques a whole new set of tools have emerged and common for them all are that they make inferences based on the population genetics of the pathogen. The basic assumption is that genetic relations between pathogen subtypes are indicators of host association or transmission pathways. These methods are also based on a comparison of subtypes from different sources and humans, but additionally take into account the genetic relatedness among the subtypes i.e. how closely are they related and how they may have evolved from each other. Some of these methods directly provide attribution estimates, where a number or proportion of human cases is attributed to specific source. These include the Bayesian clustering algorithm STRUCTURE (Pritchard et al., 2000) and the asymmetric island model (Wilson et al., 2008). Other methods are based on clustering techniques that visualise the relatedness of bacterial subtypes using some graphical representation, for example the Minimum Spanning Trees (Feil et al., 2004; Spratt et al., 2004). Although such tools do not result in risk estimates, they still provide an increased insight into the population dynamics of a pathogen and can support the conclusions drawn from more mathematical models.

Particularly for source attribution of human campylobacteriosis, population genetic methods based on MLST data have received increased attention in recent years, as it has been possible to identify some degree of host association between certain sequence types (ST) and a particular host reservoir despite the weakly clonal population structure of this pathogen (Dingle et al., 2001; McCarthy et al., 2007). The mathematical modelling approaches applied for attributing human campylobacteriosis using MLST data include the STRUCTURE model and the Asymmetric Island model (Wilson et al., 2008; Sheppard et al., 2009; Strachan et al., 2009; Mullner, 2009; Mughini-Gras et al., 2012; Boysen et al., 2013). In brief, the assumption is that the animal and environmental reservoirs of *Campylobacter* are separate populations within which the bacteria evolve through mutation and horizontal gene transfer (recombination), and between which genes may flow (migrate). Based on the estimated amount of mutation, recombination and migration, each human case is assigned probabilistically to the source populations. From these individual probabilities, the total amount of human disease attributable to each source is estimated. These techniques have so far primarily been used in attribution studies of *Campylobacter*, but it is expected that they will also be applicable to other zoonotic pathogens such as *Salmonella*, *L. monocytogenes* and STEC as sequence-based typing methods become more widely used and experiences in identifying the appropriate discriminatory level are increased.

5.1.2.3. Comparison of principles between the two types of models

Genetic diverse pathogens and/or data, based on subtyping with a high discriminatory level are a challenge for the frequency-based models, as the apportioning of cases is based on exact match between subtypes in sources and humans. The possibility of cases being attributed to a category of 'unknown' allows for the human cases not to fit within the included sources; rendering the possibility of reservoirs not being represented. This approach may therefore be considered more conservative than the population genetic models. In other words, no sources will be incriminated unless identical subtypes are found in both humans and the source question, but the approach may potentially lead to a large group of human cases that cannot be attributed to any source.

The population genetic models have the advantage of considering the relatedness of isolates, taking estimates of recombination, mutation and migration into account. Considering the substantial genetic variation of for instance *Campylobacter*, this may be considered a strength. The model estimates the relative probability of each human case to belong to each source included, and estimates the total proportion of cases attributable to these sources (Wilson et al., 2008). The models do not operate with an 'unknown' source category, meaning that there is a potential risk for human cases being attributed it to non-responsible sources. Inclusion of data from all sources with human health significance is therefore critical.

5.2. Optimal data needs

Source attribution relies on data collected through an integrated surveillance i.e. a surveillance including data from both humans and ideally all putative sources. Integrated surveillance also implies that samples from each source (and humans) are collected through harmonised surveillance programs so that the resulting data can be used to estimate prevalences that are comparable between regions/countries and over time. In addition and as already mentioned in chapter 4, the isolates from humans, animals and food should be characterised using the same typing methods relying on internationally standardised nomenclatures and protocols. The typing data should be accompanied by relevant epidemiological information needed to analyse and interpret the data. Data required for source attribution analysis often originates from different registries/databases (e.g. laboratory databases, central husbandry registries or patient registries). Unique identifiers should therefore be agreed so that the data from the different databases can be merged appropriately.

Besides the typing-related human, animal and food data, inclusion of consumption data to weigh the contribution from the different sources in the frequency-based source attribution models may improve the quality of the results and is particularly useful for interpreting the results e.g. assessing whether a certain reduction in the number of cases associated with a particular source is a result of risk management actions or due to reduced consumption (EFSA, 2008). For multinational models e.g. at

the EU level, such data are considered to be essential because of the extensive trade of foods between MSs (Hald et al., 2012). Ideally, the models should employ consumption data of the specified food sources. National consumption data do not generally include detailed information of e.g. the origin of the food (i.e. the country in which the food where produced). Therefore, an approximation is often used, where the amount available for consumption produced in a country is estimated as the production minus export. In multinational models, the amount of food imported to one MS from another MS should also be included to consider trade between MSs. Data on production and trade of animal-derived food sources can be estimated based on the statistics reported to EUROSTAT, but requires a thorough data validation and some subjective decisions taken by the modeller. This is because even though the European Community legislation ensures that the statistics provided by the MSs are based on legal texts and on harmonized definitions and procedures, the quality of the trade data reveals major and persistent inconsistencies in the various MSs intra-EU trade statistics (EFSA Panel on Biological Hazards, 2011b). The use of EUROSTAT data is, therefore, nether ideal or straightforward and initiatives to improve the data reported to the EFSA Comprehensive European Food Consumption Database, so that these can be used for source attribution, is recommended.

5.3. Current experiences in source attribution employing molecular sub-typing methods

The application of data based on phenotyping has the advantage that the same methods often are used for surveillance of animals, food and humans. Results are therefore usually more readily available than molecular data, and the results are comparable between laboratories, which are in contrast to, for instance, MLVA and PFGE, which can be more difficult to standardise. This means that the data usually represents better the geographic and temporal relations between isolates from humans and possible sources, and data from more sources can often be considered. Another advantage is that a combination of different phenotypes can be used to define a subtype (e.g. serotyping, phage typing and antimicrobial resistance patterns) to be included in a source attribution model.

The most obvious weakness of the phenotypic methods is that the discriminatory level may be too low, particularly if only using one kind of phenotypic data (e.g., serotypes of *Salmonella*), or that the existing phenotypes do not cluster genetically related isolates (e.g. serotypes of *Campylobacte*r, which often seems to be related to horizontal gene transfer). The latter indicates that in order for phenotypes to be useful, the genetic and phenotypic relationships among different lineages needs to be established. With regard to choice of model phenotypic data can only be applied in the frequency-matched models.

Experiences with using molecular methods for source attribution, as defined in this opinion, are quite limited. The most investigated pathogen is *Campylobacter* for which several studies has been conducted using MLST typing. Although some of the most common MLST types are found in both humans and several animal food sources, the results have been able to suggest the most important reservoirs for human infections (Wilson et al., 2008; Sheppard et al., 2009; Strachan et al., 2009; Mullner, 2009; Boysen et al., 2013). Still, current MLST typing fails to differentiate between animal species within the ruminant reservoir (i.e. cattle, sheep and goats), and for some major STs, discrimination between the chicken and ruminant reservoirs is not possible.

One study included an additional discriminating attribute, the sequenced *flaA* gene, but this inclusion did not cause considerable changes in the results (Boysen et al., 2013). In a study by Mughini-Gras et al. (2012), a Dutch case-control study of human sporadic cases of *Campylobacter* was re-analysed by subdividing the dataset by MLST ST. The study showed that combining epidemiological and molecular typing data improved the identification of risk factors, and showed that MLST-based source attribution for human campylobacteriosis makes epidemiological sense.

Campylobacter is a very diverse group of bacteria, so methods focusing on a set of well conserved genes, such as MLST have several advantages. The technique is unaffected by changes in the gene order along the chromosome, which can be altered as a result of intragenomic recombination, and MLST typing has proven to be sufficiently discriminatory to identify associations between certain MLST types and animal reservoirs. Furthermore, the method is easily reproduced in different

laboratories. In contrast, MLST typing of *Salmonella* is only capable of classification at the species level and sometimes also at the subspecies or serovar level (Leekitcharoenphon et al., 2012), but will not be sufficiently discriminatory for the purpose of source attribution. Thus for *Salmonella*, more or other parts of the genome content should be considered. A recent attempt to apply WGS and Bayesian analysis to a limited collection of isolates of *S*. Typhimurium DT 104 has been described by Mather et al., 2013, but the conclusions from this study were highly influenced by the choices made in selecting strains for analysis and furthermore, were reached without sufficient consideration of existing epidemiological knowledge.

It has not been possible to find any formal source attribution studies of STEC using molecular methods. Many studies have shown that ruminants are a major reservoir of STEC infections in humans, but whether molecular methods can be used to further disentangle the transmission pathways needs further research. A study on ESBL *E. coli* from Germany (Valentin et al., 2013) applied an adapted version of the Danish frequency matched model with the purpose to identify possible animal reservoirs of ESBL infections in humans. Subtypes were defined based on occurrence of ESBL-genes and the phylogenetic group. Preliminary results indicated that many human cases cannot be explained by the animal sources included, but the authors also concluded that the typing information considered to define a subtype was not sufficiently discriminatory.

A source attribution study of human listeriosis has been published by Little et al. (2010). The study used an adapted version of the Danish model and included a combination of *L. monocytogenes* serotypes and AFLP types. The attribution estimates showed very wide credibility intervals, which the authors discussed could be a result of the complicated epidemiology of listeriosis, particularly the poorly understood dose response relationship. However, looking at the data presented in the paper, it t seems also that the subtyping approach could have benefited from a higher resolution, as many of the most common subtypes in humans were also found in several food sources. Another limitation of applying typing approaches to *Listeria* is the fact that this pathogen is primarily found as a contaminant of the processing environment and does not as such have direct animal reservoirs, although resident processing equipment contaminants are likely to have originally come from the food being processed. For future studies, a redefinition of 'reservoir' for *Listeria* as e.g. a specific processing plant may make more epidemiological sense.

Finally, one of the most important lessons learned when reviewing available source attribution studies is the requirement for including isolates from humans and all potential major sources that are related in time and space. Because of lack of relevant data, some studies have used surrogate data, including data from different geographic regions and/or time periods, whereas other studies have simply not considered the relative occurrence of specific phenotypes or sources that are otherwise known to play an important role (e.g. Mather et al., 2013). This may seriously bias the model results (Smid et al., 2013). So, besides appreciating the population diversity and structure of the pathogen in question, the data and results should always be interpreted in the right epidemiological context. This means that additional information relating to the data, such as time of sampling, origin of the sample, is of paramount importance in order to draw conclusions and interpret the attribution results.

5.4. Potential use of new technologies

The clear advantage of molecular methods compared to phenotypic methods is that the former in general have a much higher discriminatory level. In fact, the application of WGS should in theory be able to provide us with all we need to know about a certain bacterial isolate. There will be major challenges in how to analyse and interpret this enormous amount of data. Specifically for source attribution, important research should be focused upon ways to define meaningful subtypes that can be used as input for the mathematical attribution models. This problem refers back to the issue discussed above about defining an appropriate level of discrimination, i.e. defining the isolates that can be regarded as epidemiological related. It is also to be expected that this discriminatory level will vary depending on, among other factors, the clonality of the pathogen analysed.

A complicating factor which is also discussed in chapter 2 is the fact that most bacteria do not conform strictly to clonal models, but exhibit variable rates of horizontal gene transfer. Such gene transfer may include elements such as virulence genes or genes that code for antimicrobial resistance elements that distort the genetic relationships among isolates, but may be important to consider in source attribution studies in order to identify epidemiologically related and non-epidemiologically related isolates. As examples, in *Salmonella* the evolution of virulence is largely driven by horizontal gene transfer and antimicrobial resistance genes are often located on mobile genetic elements such as plasmids. Molecular methods that only work at the chromosomal level, e.g. MLST, will not include variation in such genes. It is therefore critical that the molecular method chosen includes the appropriate information on genetic variation of the pathogen investigated and that this variation can also be considered by the attribution models. This is a challenge for the currently existing population genetic methods, which are mainly considering genetic relationships. Subtypes applied in frequency-matched attribution methods should obviously also include appropriate information on genetic variation. Since these methods do not consider the genetic relationship between isolates, and the subtypes, in principle, can be defined by a combination of all kinds of phenotypic or genotypic information, the data analysis is more straightforward.

As already described, MLST has been used with success for source attribution of *Campylobacter*, whereas the low discriminatory capability is inappropriate for *Salmonella*. However, as shown for *Campylobacter* in recent studies (e.g. Penny et al., 2013; Boysen et al., 2013), the use of additional markers (e.g. *fla*A, *por*A and *gyr*A) besides the traditional housekeeping genes can increase the discriminatory power, which could be useful for typing of *Salmonella* and other food-borne pathogens for source attribution. Other methods include PFGE and MLVA, which are most commonly applied for surveillance, particular outbreak investigations. These methods certainly have their benefits, but they are labour-intensive and difficult to standardise, and their usefulness for source attribution remains unexplored. With the high-throughput WGS techniques, a high level of standardisation and consequently meaningful comparison of results between technologies is expected. Here the challenge will be to select genes for determining appropriate subtypes. Such subtypes could include one or more of the housekeeping genes, but they could also be based on a whole new set of genes identified through WGS.

In the long term, WGS may be able to provide the knowledge/information that is needed to quantify the difference between various subtypes/strains with regard to causing human illness and thereby assist with the characterisation of 'pathotypes'. This discussion is particular relevant for STEC, where MLST combined with determination of virulence genes provide better insight into identifying food/animal strains relevant for human disease than MLST alone (Hauser et a., 2013; Ji et al., 2010), and virulence genes are considered important to include in subtypes used for STEC source attribution. A source attribution approach that is able to consider both genetic and functional relationship between isolates would be very useful, especially if the functional traits relate to factors important for human infections such as virulence, antimicrobial resistance and survivability (e.g. acid tolerance).

5.5. Concluding remarks on the use of molecular typing methods in food-borne source attribution

The following are concluding remarks on the use of molecular typing methods for food-borne source attribution:

- A major challenge of using data generated from molecular typing methods in source attribution models, in particular WGS data, will be to define meaningful subtypes providing an appropriate level of discrimination for source attribution. A high level of discrimination is not necessarily the best option. The applied method has to allow for some genetic diversity between isolates from human and animal/food sources, but only to the degree so that it can still be assumed that they originate from the same source.

- Independent of the choice of molecular typing method and approach for source attribution, it is important that the data included from human and potential sources are related in time and space. Source attribution analysis is, therefore, facilitated by integrated surveillance providing a collection of isolates from all (major) sources that should, to the extent possible, represent what the human population is exposed to.

- Reproducibility, standardisation and high degree of typeability are additional essential criteria for the chosen typing method, and the typing data should be accompanied by relevant epidemiological information needed to analyse and interpret the data.

- The microbial subtyping approach for source attribution is limited to pathogens that are heterogeneously distributed among the reservoirs. This makes it appropriate for pathogens, which are clonally distributed and present at least some host-associated subtypes.

- The most commonly used source attribution models using subtyping approaches can be divided into two categories: (a) the frequency-matching models and (b) the population genetic models.

- Frequency-matching attribution models for *Salmonella* have only used phenotypic data, whereas molecular methods have also been applied for *Campylobacter* (MLST). For attribution of *Listeria monocytogenes* a combination of phenotyping (serotyping) and molecular typing (AFLP) has been used. For *Salmonella*, the usefulness of molecular methods for source attribution (e.g. MLVA and WGS-based methods), perhaps in combination with phenotypic methods, still needs to be explored.

- Population genetic models (e.g. STRUCTURE and the Asymmetric Island model) have so far primarily been used in attribution studies of *Campylobacter*. It is expected that they will also be applicable to other zoonotic pathogens such as *Salmonella*, *Listeria monocytogenes* and STEC as sequence-based typing methods become more widely used and experiences in identifying the appropriate discriminatory level are increasing.

- Adaptations of existing source attribution models and/or development of new models (and software) that are able to explore and compare the use of many different combinations of subtypes and functional genetic traits is recommended.

6. The use of molecular typing methods in the early identification of food-borne organisms with epidemic potential and their integration in risk assessment

6.1. Data needs for the identification of food-borne organisms with epidemic potential

The ultimate goal of molecular typing for identification of food-borne organisms with epidemic potential is to predict virulence characteristics of putative pathogens from genomic information. Such predictions can also be useful for more precise and targeted 'predictive' hazard identification within the risk assessment process. The identification of such microbiological hazards at an early stage can thus prevent further spread of strains with high virulence and/or epidemic potential. Furthermore, it may be possible to focus control measures and/or microbiological criteria to (the most) pathogenic variants of bacterial species, for example highly virulent variants of STEC (EFSA BIOHAZ Panel, 2013). The information can also be useful to better incorporate variability between bacterial strains in risk assessment models, particularly with regard to their survival in the food chain and in dose-response relationships. An ideal typing method would be able to both compare a newly-emerged organism with existing strains and to identify relevant characteristics in such emerged organisms that as yet have no history of enhanced epidemic capacity or virulence.

Proactive measures rely on waiting until a food-borne epidemic has been identified, or a highly virulent strain has emerged and has been reported in certain region or regions. Such measures may then encourage other regions to apply intervention measures, such as trade restrictions, heat treatment of contaminated foods and actions in infected livestock identified as reservoirs or vehicles of infection. Hopefully such measures would assist in reducing the introduction or further spread of the organism, or eliminating it from a food vehicle (Davies et al., 2013). If reliable genetic markers for identifying virulence-associated characteristics of potentially hazardous strains could be identified, then the appearance of such strains could raise alerts at an early stage, thus leading to the implementation of control measures that could prevent or reduce the impact of an epidemic. Organisms with epidemic potential possess characteristics that can enhance their dissemination, virulence or persistence, either in the host or in ecological niches. There are numerous examples of dissemination of strains due to epidemiological opportunities, such as the occurrence of an organism at the top of a food animal breeding pyramid or in a major animal feed ingredient. Such occurrences may lead to the wide dissemination of strains in food animals. Furthermore, the ability of some organisms to evade, subvert or delay the host immune response can lead to more rapid spread. At the same time, properties such as resistance to antimicrobials, disinfectants, heavy metals or adverse environmental conditions such as heat or desiccation can supply a competitive advantage allowing the proliferation of a pathogen (Ma et al., 2013).

In (quantitative) microbial risk assessment, whether the hazard considered is a genus or species (e.g. *Salmonella enterica*), serovar (e.g. *S.* Typhimurium) or a subtype (e.g. *S.* Typhimurium DT 104) is defined by the resolution in the data available. In most of the cases, depending on the resolution of epidemiological and microbiological data, the primary assumption is that all strains within a genus, a species, or a serotype behave similarly, both in the food chain and in human hosts. Even if data at a high level of resolution are available, the assumption of similar behaviour may not be justified. For example, Berk et al. (2005) demonstrated strong variability in acid resistance between isolates of *S.* Typhimurium DT 104 from different origins. All isolates with high acid resistance were obtained from humans. Similar differences between strains, even when subtyped to a high level of detail, will exist with respect to many other relevant properties. Thus, typing should ideally be based on characteristics that are directly relevant for microbial behaviour along the food chain and in humans.

To identify strains with epidemic potential in the absence of comparability with existing epidemic strains the gene content of such strains should be assayed in order to predict the phenotype and how such genes guide the interaction of the strains with their environment. This is a highly complex matter, in which detailed understanding at several levels is required:

- **Genetic information** based on molecular characterisation methods that index variation relevant to the human host and food chain.

- **Expression information**, including transcription of DNA into RNA and proteins, in interaction with the (host) environment.

- **Ecology of pathogens** (e.g. host-specificity, survival in food chains and other environmental niches, colonisation of humans, interaction with the human microbiome and immune system).

- **Pathogenicity**, this is the potential to initiate disease processes in human hosts.

Current studies mainly aim at directly linking genetic information to pathogenicity, as measured by a wide range of *in vitro* and *in vivo* experiments aiming to identify virulence factors. Such studies are usually undertaken in mice. Even when direct extrapolation of results to food animals and humans is difficult, the mouse model can provide useful indicative information, particularly when transgenic mice are used. Alternatively and most importantly, by linking typing data to comprehensive epidemiological data, the pathogenic potential of subtypes and their complement of virulence and stress-response genes can be evaluated. In some cases such studies have identified associations between single virulence factors (e.g. specific Stx2 subtypes and the number of gene copies) and pathogenic potential. In other cases, several virulence factors may be involved, some of which can be clearly delineated and some which are less clearly defined. Often the search for virulence factors has not been productive and the disease process is the result of a complex interplay between the infecting organism and host defence mechanisms, or clonal expansion within a favourable epidemiological niche (e.g. *S.* Enteritidis in the chicken breeding and production pyramid or *S.* Typhimurium in pig breeding and distribution). The multi-factorial inter-relationships required for a pathogen to become a 'potential' epidemic strain can be described by employing a modified version of the classical pathogen-host-environment relationship, as shown below in Figure 5.

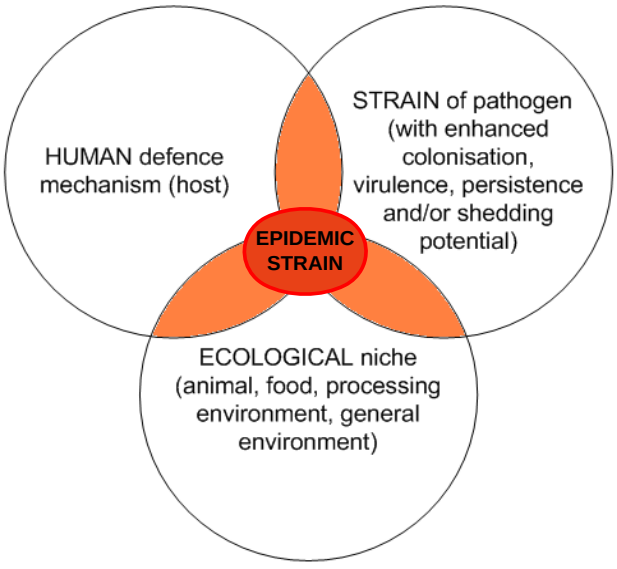

Figure 5: Interaction of pathogen-host-ecological factors that influence food-borne disease epidemics.

This multi-factorial framework drives the complexity behind elucidating which strains may have food-borne epidemic potential, and should be taken into account when considering genomics and molecular typing methods as tools for 'predictive' hazard identification. As yet identification of a single genetic marker or combination of markers that would qualify a strain with such epidemic potential is not possible, or at least it has not yet been done in a prospective or predictive manner.

Nevertheless, and with regard to the four pathogens considered in this Opinion, the following are examples of considerations that should be taken into account when exploring the use of molecular typing methods for prospectively identifying strains from food-borne organisms with epidemic potential:

- *Salmonella* spp.

  Ten years after publication of the first *Salmonella* genome still thousands of genes remain of hypothetical or unknown function, and the basis of colonisation of reservoir hosts is still poorly defined. Genomic diversity across bacterial strains is largely shaped by gain of functions via horizontal gene transfer. The chromosomal acquired genomic islands that encode virulence genes are referred to as pathogenicity islands. In *S. enterica*, 21 *Salmonella* Pathogenicity Islands (SPIs) have so far been identified, in addition to the *Salmonella* genomic island 1 (SGI-1) and the high-pathogenicity island (HPI). SPIs are considered to be 'quantum leaps' in the evolution of *Salmonella*, playing a fundamental role in pathogenesis and host specificity/adaptation. While certain SPIs (such as SPI-1 and SPI-2) have been studied in depth, other SPIs have been identified more recently and much less is known about their distribution across *Salmonella* serovars and the role they play in disease. *Salmonella* has evolved a complex functional interface with host cells largely determined by two type III secretion systems (T3SS). These are encoded within SPI1, which through the delivery of bacterial effector proteins modulates a variety of cellular processes that facilitate uptake into, and replication within, epithelial and phagocytic cells (Hannemann et al., 2013). Further virulence traits, such as the pSLT virulence plasmid, adhesins, flagellae, ion transporters, superoxide dismutase and biofilm-related proteins, also contribute to success within the host (Ibarra and Steele-Mortimer, 2009). Several regulatory mechanisms which synchronize all these elements in order to guarantee bacterial survival have been described (Suez et al., 2013).

  Only a limited number of *Salmonella* serovars within subspecies *enterica* carry a large, low-copy-number plasmid that contains virulence genes. Virulence plasmids are required to trigger systemic disease but their involvement in intestinal colonisation is unclear. *Salmonella* virulence plasmids are heterogeneous in size (50–90 kb), but all share a 7.8 kb region, spv, required for bacterial multiplication in the reticuloendothelial system. The *spv* region contains three genes required for the virulence phenotype including SpvB which exhibits a cytotoxic effect on host cells and is required for delayed cell death by apoptosis following intracellular infection. Strains isolated from systemic infections of immune compromised patients, particularly HIV patients, usually carry the *spv* locus, suggesting that CD4 T cells are required to control disease due to *Salmonella* that are *spv*-positive (Guiney and Fierer, 2011). Other loci of the plasmid, such as the fimbrial operon *pef*, the conjugal transfer gene *traT* and the *rck* and *rsk* loci, may play a role in other stages of the infection process. The virulence plasmid of *S.* Typhimurium is self-transmissible but virulence plasmids from other serovars, such as *S.* Enteritidis and *S.* Choleraesuis, carry incomplete *tra* operons. The presence of virulence plasmids in host-adapted serovars suggests that virulence plasmid acquisition may have expanded the host range of *Salmonella* (Fàbrega and Vila, 2013).

  Multiple antibiotic resistance has been suggested as one of the reasons for emergence of epidemic *S.* Typhimurium strains such as those represented by DT104 and monophasic U302 and DT193 variants (Hall, 2010). Multiple resistance or resistance to high priority therapeutic antibiotics is also important in its own right and the existence of hybrid resistance/virulence plasmids promotes the spread of virulent strains in situations where there is antibiotic selection pressure (Beceiro et al., 2013). The importance of the identification of resistance genes of relevance to public health by molecular methods has recently been highlighted by the recognition of the emergence of resistance to carbapenems in *E. coli* and *Salmonella* from food-production animals in Germany (Fischer et al., 2013a,b). Plasmid typing analysis by identification of replicons associated with predominant conjugative plasmids of *Enterobacteriaceae* has been extensively used for following the epidemic spread of drug

resistance-encoding plasmids (Carattoli, 2013; Johnson et al., 2012; Garcia-Fernandez et al., 2008; Hopkins et al., 2006). Not all resistance or virulence-related genes that are identified by molecular methods are consistently expressed, in order to enhance the fitness of the organism (Humphrey et al., 2012), but identification of the genes can serve as a trigger for further phenotypic investigations if necessary.

Human infection with non-typhoidal *Salmonella* serovars occasionally results in invasive systemic disease and bacteremia, with certain serovars such as *S*. Choleraesuis and *S*. Dublin, which are also invasive in their preferred hosts, being proportionately more likely to result in systemic disease. Comparative genomic hybridization using a *Salmonella enterica* microarray has revealed a common core of 3233 genes present in invasive strains, which include the *Salmonella* pathogenicity islands 1–5, 9, 13, 14; five fimbrial operons (*bcf, csg, stb, sth, sti*); three colonization factors (*misL, bapA, sinH*); and the invasion gene, *pagN*. Additional novel genomic islets; various *Salmonella* virulence factors; and several typhoid-associated virulence genes (tcfA, cdtB, hlyE, taiA, STY1413, STY1360), are also more widely distributed amongst *Salmonella* serovars than previously thought (Suez et al., 2013).

Most studies of the role of *Salmonella* genes *in vivo* have focused on systemic virulence in murine typhoid models, and the genetic basis of intestinal persistence and thus zoonotic transmission in food animals has had little attention. Transposon-directed insertion-site sequencing is beginning to elucidate the complexity of genetic mechanisms involved in infection but much remains to be done to identify the complex complementary patterns of genes and confirm those that are essential for pathogenicity (Chaudhuri et al., 2013; Hammarlöf, et al., 2013).

- *Campylobacter* spp.

Specific virulence mechanisms have not yet been clearly elucidated for *Campylobacter* spp. The ability to reach the intestinal tract is, in part, due to resistance to gastric acids and also to bile salts. Flagella-mediated motility, bacterial adherence to intestinal mucosa, subvasion (i.e. active and rapid migration of the pathogen into the subcellular space (van Alphen et al., 2008)) and invasion capability and the ability to produce toxins have been identified as putative virulence factors. Flagellae not only facilitate motility but also secretion of *Campylobacter* invasive antigens. In contrast to other diarrhoea-causing bacteria, no other classical virulence factors have yet been identified in *C. jejuni* but host factors leading to an acute intestinal inflammatory response appear to play a major role in pathogenesis. Different adaptation strategies are adopted depending on its current requirement, e.g. multiplication and persistence in the natural avian reservoir or environmental survival (Dasti et al., 2010). At a molecular level, recent developments in deciphering mechanisms of virulence show that *C. jejuni* is a unique pathogen. Although *C. jejuni* has a limited ability to metabolize sugars, the organism possesses a large number of enzymes involved in the biosynthesis of carbohydrates which are then incorporated into its peptidoglycan (PG), lipooligosaccharides (LOS), capsular polysaccharides (CPS) and both N- and O-linked glycoproteins. Recent studies have indicated a role of the lipooligosaccharide (LOS) of *C. jejuni* in virulence, particularly the severe neurological Guillain Barré syndrome (Ellström et al., 2013; Pike et al., 2013). LOS and CPS are also involved in the interaction between *Campylobacter* and the human innate and acquired immune system. Full definition of the genetic basis of the pathogenicity of *Campylobacter* has been limited by the lack of suitable in vivo models, but information is growing and linkage of extensive gene array and sequencing data with epidemiological and metagenomic features should facilitate identification of more definitive gene targets for risk assessment (Diaz-Sanchez et al., 2013; Marotta et al., 2013; Melo et al., 2013; Szymanski et al., 2012). The use of gene association studies comparing organisms with different phenotypes is likely to be influential in this area (Sheppard, et al., 2013b).

MLST has been very informative in associating certain *Campylobacter* clonal complexes or sequence types with different animal hosts, particularly poultry. Several multi locus sequence types (STs) that are frequently isolated from wild birds, are rarely if ever observed in humans (Griekspoor et al., 2013). This might be explained by limited exposure of humans to wild bird strains, but even then, such strains are also not usually seen in livestock. Understanding the molecular basis for the host range of *Campylobacter* STs might provide insight in the genetic determinants of human pathogenicity. Human risk seems to be primarily associated with livestock-associated STs, including those from cattle that may be found as contaminants of surface water and private water supplies. MLST results are confirmed by approaches using more genomic information, including WGS, although the latter offers more detailed insights, particularly for widespread STs (On, 2013). Initial steps to study differences in virulence factors are being made and it is likely that there will be a rapid expansion in available information in the near future, but currently it is not clear which factors are associated with the observed differences in host range of *Campylobacter* strains.

- *Listeria monocytogenes*.

  Despite numerous studies that have identified panels of virulence and stress response genes, as outlined in Chapter 2, there is still much to learn about the detailed pathogenesis of *Listeria monocytogenes*. Whole genome comparative analysis has revealed that the *L. monocytogenes* genomes are essentially syntenic, with the majority of genomic differences consisting of phage insertions, transposable elements and SNPs. These strain and serotype specific genes probably contribute to observed differences in pathogenicity, and the ability of the organisms to survive and grow in their respective environmental niches (Nelson et al., 2004). Virulence genes in *Listeria* can be detected using PCR, but strains carrying such genes are widespread in foods such as cheeses, so additional epidemiological information is needed to inform decisions on their epidemic potential (Lomonaco et al., 2012). A haemolysin, listeriolysin S, is associated with the majority of outbreak strains and can be detected by a rapid PCR test (Clayton et al., 2011). PCR has also been used in the recent identification of Tn*6188*, a transposon in *L. monocytogenes* that carries a *OacH* gene that confers tolerance to benzalkonium chloride, a quaternary ammonium compound disinfectant that is widely used in the food industry (Müller et al., 2013). Another gene product is a positive regulator of multiple virulence determinants in *L. monocytogenes* (Chakraborty et al., 1992). This complexity of virulence gene regulation has yet to be fully defined (Lobel et al., 2012; Shen et al., 2013). WGS offers a method that is both highly discriminatory and more informative than previous technologies used for characterisation of *Listeria* and identification of outbreaks (Knabel et. al. 2012), but there is a need for wider studies that link genetic and epidemiological data to fully assess the significance of potential markers of epidemicity.

- STEC.

  As discussed in Chapters 2 and 4, STEC are genetically heterogenous with a high degree of genome plasticity. As such prediction of both pathogenicity and epidemicity is difficult. There is no single or combination of marker(s) that defines the potential of a STEC strain to cause human disease, as various factors and toxins contribute to the virulence of STEC. Shiga toxin type 2 (Stx2) is more often associated with confirmed cases of human disease, and those strains producing this toxin are more frequently associated with severe illness. Strains that produce Stx2 have been suggested to be more likely to cause HUS than those that produce Stx1 alone. Detection and identification of Stx variants can be done by PCR, or PCR with subsequent sequencing (Scheutz et al., 2012). A molecular approach, utilising tests for the presence or absence of *eae*, *aaiC* or *aggR* genes additional to the presence of s*tx* genes, has recently been proposed (EFSA BIOHAZ Panel, 2013), but needs to be verified with well-characterised isolates from cases of human infection and from food-producing animals and foods.

The urgent investigation of the large epidemic of STEC O104:H4 in the EU in 2011 (see Chapter 4) demonstrated the value of combined serotyping and genotypic analysis coupled with the rapid communication of results through a variety of media. In view of the widespread utilisation of WGS, this investigation should be taken as a model for the identification of highly virulent STEC strains with epidemic potential. More recent advanced molecular investigations have suggested that incursion of *Stx*2a carrying bacteriophages originating from STEC in cattle into O104 strains is likely to have been responsible for the genesis of STEC O104:H4 (Beutin et al., 2013).

Comparing the short descriptions given above of current insights and challenges in the identification of factors that are associated with the ability of a bacterial strain to cause human infection following transmission through the food chain, it is clear that no general approach can currently be developed. Furthermore, combinations of genes that predict behaviour in the food chain need to be defined for each species. This is a considerable challenge, but of great importance, both scientifically and in order to better inform risk management. Knowledge of the associations of gene content with persistence on particular ecological niches (e.g. processing equipment and in growth in certain food products) will also be relevant (Crerar et al., 2011). Strains with limited public health relevance may also be identified with more precision in order to make more efficient and targeted control decisions.

Genomic studies on the pathogen molecular characteristics, on epigenetic factors, on host-pathogen interaction and on factors related to the survival and growth of pathogens in the food chain have not yet been integrated in a single framework. This might enable a comprehensive evaluation of the pathogenic potential of a new strain identified in food animals. Such integration is indeed a formidable task, but many elements are already available or are becoming available so that, in principle, multidisciplinary research should be able to address this task.

Sophisticated analysis of complex data is also needed to assess epidemiological factors and related risks together with the genetic findings (Chattopadhyay et al, 2013). This requires fully representative panels of strains which have not mutated during storage and have good quality linked epidemiological and clinical data (Gardy et al., 2011).

In practical terms, organisms could be ranked following a probability risk matrix, that would consider epidemiological and genetic data combined with data from *in vitro* and *in vivo* experiments that has been anchored by the behaviour of known pathogenic strains. Increasingly, *in silico* experiments will inform such type of ranking. All these retrospective and prospective data can be used for developing prediction criteria for epidemic potential and virulence of newly emerging strains before such potential gets realised, ideally in time to identify the source and limit the dissemination of future major pathogens. If this is possible to achieve with a satisfactory level of confidence, it should be a valuable tool to assist in decision making on intervention strategies (Krishnaswamia et al., 2013).

6.2. Future perspectives and potential use of new technologies

With the advent of high throughput sequencing technology, it should be possible to assay thousands of strains in order to identify multiple markers. This will help in identifying combinations that have statistical associations with epidemic potential or disease severity in a similar way to what has been done in relation to genetic risk for certain cancers in people (Roukos, 2013; Chaudhuri et al., 2013). WGS using high throughput methodology offers an opportunity to derive genetic information that would be equivalent to all that is gathered by the multiplicity of other methods (Struelens and Brisse, 2013) and whole genomic mapping attempts to bridge older and WGS methodologies (Miller, 2013). For example, studies on *Campylobacter* in New Zealand have already shown that sequence data can be fitted to evolutionary and epidemiological models to gain new insights into pathogen evolution, the nature of associations between strains of pathogens and host species, and aspects of between-host transmission. With the advent of newer sequencing technologies and the availability of rapid, high-coverage genome sequence data, such techniques may be extended and refined within the emerging discipline of genomic epidemiology (Muellner at al., 2013).

A recent publication from Monk et al. (2013) reports the construction of genome scale models (GEM) of metabolism for 55 fully sequenced *E. coli* and *Shigella* strains. The core genome consisted of 965 metabolic genes, whereas the pan genome consisted of 1 460 metabolic genes. The strain set included commensals, intestinal pathogenic *E. coli* (InPec) and extraintestinal pathogenic *E. coli* (ExPec). Significant differences in the ability to catabolise specific compounds were found between these groups. For example, InPec strains could be differentiated from commensal strains by the inability to catabolise 5 nutrients, including fructoselysine and psicoselysine, even though 11% of commensal strains were not predicted to do so. The accuracy of GEM predictions was 80% when compared to growth of 11 strains in minimal media on discriminating nutrients. When data become available, it may be possible to investigate whether differentiation between pathogenic and non-pathogenic strains of food-borne organisms is possible at an even higher level of resolution, e.g. discriminating pathogenic from non-pathogenic STEC strains.

Even with the use of new technologies, one of the biggest challenges is to combine data on genome sequences with those on interactions of the bacteria with their environment and ultimately with the human host. This will require multidisciplinary research teams to be formed with adequate focus and resources, which will enhance 'predictive' hazard identification by identifying those factors that contribute to discriminating candidate strains with epidemic potential from other strains. Further, WGS analysis alone may not be sufficient, even though providing high resolution data, as there may be other regulatory mechanisms governing the level of expression of genes under different conditions. The identification of the genetic basis of such molecular switches is advancing quickly.

If an epidemic strain has already emerged in a certain region it can be rapidly characterised by a variety of phenotypic and molecular typing methods. This can still serve to promptly identify the occurrence of such strains in other regions so to be subjected to early controls and risk management measures. WGS may provide fast and accurate typing data for comparing strains with very high resolution. Although there are differences between bacterial species, the principle of assessing the gene content in relation to epidemiological fitness as a means to assess risk potential that has been used for the four organisms considered in this opinion should be applicable to any micro-organism, e.g. *Yersinia enterocolitica* (Dhar and Verdi (2013) or *Mycobacterium tuberculosis* (Waddell et al., 2013).

Consideration should also be made regarding how WGS can be used to guide selection of representative isolates for archiving for future studies, and how epidemiologically important isolates can still be obtained when bacteriological culture is replaced by in-situ molecular diagnostic techniques.

6.3. Concluding remarks on use of molecular typing methods in the early identification of food-borne organisms with epidemic potential and their integration in risk assessment

New developments have enabled information about the gene content of a strain to be linked with certain phenotypic characteristics. Such knowledge may enable comparison with strains seen in earlier epidemics or outbreaks, or which exhibit high disease severity. As yet been prediction of the food-borne public health risk of emerging strains has not been possible. This is because of the complexity of gene and host-pathogen interactions, together with ecological and opportunist events that may result in the emergence of a strain with epidemic potential. With increasing technological and analytical capacity, combinations of genotypic elements may be more precisely linked with the predictive likelihood of a strain becoming a candidate for food-borne epidemic. This will require extensive studies of archived strains that have associated good quality epidemiological data to identify characteristics of organisms that are likely to become emergent epidemic strains. If such characteristics can be identified, assessment of the probability of serious consequences of the occurrence of such organisms in relation to the situation of their occurrence may be possible.

Many factors need to be considered in order to assess and manage the risk in the most efficient and effective way and large scale research is urgently required to fully understand this complexity as soon as possible in order to derive benefits for pathogen control initiatives.

In summary, the following are concluding remarks on the use of molecular typing methods in the early identification of food-borne organisms with epidemic potential and their integration in risk assessment:

- The epidemic potential of a food-borne strain within a bacterial species, or even within a subtype varies considerably, and is a function of its inherent genetic characteristics and their expression combined with ecological factors including the opportunities to spread in the food chain.

- Prediction of the public health risk and epidemic potential of emerging strains of food-borne pathogens has not yet been possible. Nevertheless, if an epidemic strain has already emerged in a certain region such a strain can be rapidly characterised employing current molecular typing methods and thus serve to identify the occurrence of such strains in other regions for risk management purposes.

- High throughput WGS technologies offer new opportunities to characterise bacterial strains in great detail. The genetic information that these technologies provide will however need to be considered together with gene expression, host and ecological factors, including the epidemiological opportunities to spread in the food chain.

- Although there are differences between bacterial species, the principle of assessing the gene content in relation to fitness as a means to assess risk potential that has been used for the four organisms considered in this opinion should be applicable to any bacteria.

- Multidisciplinary and integrated research programs are needed to develop and validate the use of detailed genetic information for 'predictive' hazard identification, accounting for gene expression and how this affects the fate of pathogens in the food chain and their interaction with human and animal hosts.

CONCLUSIONS AND RECOMMENDATIONS

CONCLUSIONS

General Conclusions

- All bacteria are subject to genetic change (e.g. in response to environmental stress and human interventions such as antimicrobial or heavy metal use or vaccination), sometimes by mutation but more often by acquisition or loss of genetic elements. These changes can be followed by clonal expansion in the case of biologically successful organisms.

- Ongoing evolution driven by genetic change and selection has given rise to highly adaptable organisms that are able to exploit and expand into novel niches and extend their host range. Such evolution may also be linked to the emergence of various 'epidemic' strains of pathogens, such as *Salmonella*, in combination with other biological factors and epidemiological opportunities for dissemination.

- The molecular characteristics of organisms provide markers for investigation of outbreaks, attribution studies, and assessment of potential virulence or epidemic potential.

- Even with high-resolution molecular approaches, up to and including whole genome sequencing (WGS) analysis, it is not possible to establish how closely two isolates are related without an appreciation of the structure and diversity of the bacterial population in question.

- To properly evaluate typing methodologies, data from strain characterisation should be linked with epidemiological metadata and the strain selection must be unbiased and statistically representative of the population to be assessed.

- International harmonisation of molecular characterisation outputs by means of standardisation or appropriate quality control procedures is essential. This includes controlling the accuracy of production of DNA sequences from WGS and the further interpretations of annotation pipelines.

Reply to the Terms of Reference

Term of Reference 1: ***Review information on current and prospective (e.g. WGS) molecular identification and sub-typing methods for food-borne pathogens (e.g. Salmonella, Campylobacter, Shiga toxin-producing Escherichia coli (STEC) and Listeria) in terms of discriminatory capability, reproducibility, and capability for international harmonisation.***

- Molecular typing methods should ideally provide appropriate discriminatory power, reproducibility, capability for international harmonisation and reduced handling of and exposure to pathogens in the laboratories. No current typing method, whether phenotypic or molecular, complies with all these expectations.

- Several methods are often used in combination in order to obtain the resolution needed. The methods applied depend on the pathogen and on the application sought. These methods have proven track records of use, and for some of them (e.g. Multi locus sequence typing (MLST), Pulsed-field gel electrophoresis (PFGE)) extensive databases of valuable typing data have been collected.

- Methods based on WGS can replace and are increasingly replacing the numerous different methodologies currently in use in human and veterinary reference laboratories, and the same methods can be used for all organisms. An essential precondition is the availability of quality

control methods, to ensure the reliability and consistency of molecular data generated, coupled with high quality bioinformatics support for the analysis of the data generated.

- Regarding WGS, limited knowledge is available in relation to the technical errors that occur during sequencing and analysis and on the effect of genetic drift in the different bacterial populations over time, which may complicate the interpretation of results.

Term of Reference 2: ***Review the appropriateness of use of the different food-borne pathogen sub-typing methodologies (including data analysis methods) for outbreak investigation, attribution modelling and the potential for early identification of organisms with future epidemic potential.***

- Detection of outbreaks and their investigation in real-time would be enhanced by the generation of fully comparable molecular typing data from human, veterinary and food laboratories prior to submission to a central or connected databases.

- Some molecular typing methods (e.g. MLST, PFGE, Multi locus variable tandem repeat analysis (MLVA)) have been harmonised to a greater or lesser extent for the purpose of outbreak detection and investigation. The international development of harmonised platforms for WGS-generated data should be encouraged.

- A major challenge of using data generated from molecular typing methods in source attribution models, in particular WGS data, will be to define meaningful subtypes providing an appropriate level of discrimination for source attribution. A high level of discrimination is not necessarily the best option. The applied method has to allow for some genetic diversity between isolates from human and animal/food sources, but only to the degree so that it can still be assumed that they originate from the same source.

- Independent of the choice of molecular typing method and approach for source attribution, it is important that the data included from human and potential sources are related in time and space. Source attribution analysis is, therefore, facilitated by integrated surveillance providing a collection of isolates from all (major) sources that should, to the extent possible, represent what the human population is exposed to.

- The epidemic potential of a food-borne strain within a bacterial species, or even within a subtype varies considerably, and is a function of its inherent genetic characteristics and their expression combined with ecological factors including the opportunities to spread in the food chain.

- Prediction of the public health risk and epidemic potential of emerging strains of food-borne pathogens has not yet been possible. Nevertheless, if an epidemic strain has already emerged in a certain region such a strain can be rapidly characterised employing current molecular typing methods and thus serve to identify the occurrence of such strains in other regions for risk management purposes.

- High throughput WGS technologies offer new opportunities to characterise bacterial strains in great detail. The genetic information that these technologies provide will however need to be considered together with gene expression, host and ecological factors, including the opportunities to spread in the food chain.

- Although there are differences between bacterial species, the principle of assessing the gene content in relation to fitness as a means to assess risk potential that has been used for the four organisms considered in this opinion should be applicable to any bacteria.

RECOMMENDATIONS

- Modern molecular typing methods provide many opportunities for rapid and accurate determination of the genealogical relationships among bacterial isolates. Interpretation of the results generated by these methods for different public health applications requires this information to be placed in the context of the diversity, degree of genetic change (e.g. during storage of isolates or mutation during an outbreak and in reservoirs) and population structure of the particular pathogen in question. Therefore, large scale carefully co-ordinated studies are required to fully elucidate this.

- The development of more informative and easier to use bioinformatic tools for analysis of WGS data is needed.

- Cross-sector and international coordination of the method validation is required as a priority.

- Adaptation of existing source attribution models and/or development of new models (and software) that are able to explore and compare the use of many different combinations of subtypes and functional genetic traits is recommended.

- Multidisciplinary and integrated research programs are needed to develop and validate the use of detailed genetic information for 'predictive' hazard identification, accounting for gene expression and how this affects the fate of pathogens in the food chain and their interaction with human and animal hosts.

- Development and improvement of international initiatives with regard to harmonised platforms for sharing of data such as those promoted by Pulsenet and ECDC/EFSA should be urgently prioritised, including the integration of WGS into these platforms.

- An updated EU-wide review, similar to that carried out by EFSA in 2008, on the availability and application of molecular typing methods for food-borne pathogens including those from the public health sector is recommended.

REFERENCES

Abid M, Wimalarathna H, Mills J, Saldana L, Pang W, Richardson JF, Maiden MC and McCarthy ND, 2013. Duck liver-associated outbreak of Campylobacteriosis among humans, United Kingdom, 2011. Emerg Infect Dis, 19, 1310-1313.

Achtman M, 1996. A surfeit of YATMs? J Clin Microbiol, 34, 1870.

Achtman M and Wagner M, 2008. Microbial diversity and the genetic nature of microbial species. Nat Rev Microbiol, 6, 431-440.

Achtman M, Wain J, Weill FX, Nair S, Zhou Z, Sangal V, Krauland MG, Hale JL, Harbottle H, Uesbeck A, Dougan G, Harrison LH and Brisse S, 2012. Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. PLoS Pathog, 8, e1002776.

Alcaine SD, Sukhnanand SS, Warnick LD, Su WL, McGann P, McDonough P and Wiedmann M, 2005. Ceftiofur-resistant *Salmonella* strains isolated from dairy farms represent multiple widely distributed subtypes that evolved by independent horizontal gene transfer. Antimicrob Agents Chemother, 49, 4061-4067.

Allard MW, Luo Y, Strain E, Pettengill J, Timme R, Wang C, Li C, Keys CE, Zheng J, Stones R, Wilson MR, Musser SM and Brown EW, 2013. On the evolutionary history, population genetics and diversity among isolates of *Salmonella* Enteritidis PFGE pattern JEGX01.0004. PLoS One, 8, e55254.

Anderson RM and May RM, 1979. Population biology of infectious diseases: Part I. Nature, 280, 361-367.

Anklam KS, Kanankege KS, Gonzales TK, Kaspar CW and Dopfer D, 2012. Rapid and reliable detection of Shiga toxin-producing *Escherichia coli* by real-time multiplex PCR. J Food Prot, 75, 643-650.

Barco L, Barrucci F, Olsen JE and Ricci A, 2013. *Salmonella* source attribution based on microbial subtyping. Int J Food Microbiol, 163, 193-203.

Batz MB, Doyle MP, Morris G, Jr., Painter J, Singh R, Tauxe RV, Taylor MR and Lo Fo Wong DM, 2005. Attributing illness to food. Emerg Infect Dis, 11, 993-999.

Bäumler AJ, 1997. The record of horizontal gene transfer in *Salmonella*. Trends Microbiol, 5, 318-322.

Beceiro A, Tomas M and Bou G, 2013. Antimicrobial resistance and virulence: a successful or deleterious association in the bacterial world? Clin Microbiol Rev, 26, 185-230.

Berk PA, Jonge R, Zwietering MH, Abee T and Kieboom J, 2005. Acid resistance variability among isolates of *Salmonella enterica* serovar Typhimurium DT104. J Appl Microbiol, 99, 859-866.

Best EL, Hampton MD, Ethelberg S, Liebana E, Clifton-Hadley FA and Threlfall EJ, 2009. Drug-resistant *Salmonella* Typhimurium DT 120: use of PFGE and MLVA in a putative international outbreak investigation. Microb Drug Resist, 15, 133-138.

Best EL, Lindstedt BA, Cook A, Clifton Hadley FA, Threlfall EJ and Liebana E, 2007. Multiple-locus variable-number tandem repeat analysis of *Salmonella enterica* subsp. *enterica* serovar Typhimurium: comparison of isolates from pigs, poultry and cases of human gastroenteritis. J Appl Microbiol, 103, 565-572.

Beutin L, Hammerl JA, Reetz J and Strauch E, 2013. Shiga toxin-producing *Escherichia coli* strains from cattle as a source of the Stx2a bacteriophages present in enteroaggregative *Escherichia coli* O104:H4 strains. Int J Med Microbiol, 303, 595-602.

Bhatty M, Laverde Gomez JA and Christie PJ, 2013. The expanding bacterial type IV secretion lexicon. Res Microbiol, 164, 620-639.

Biggs PJ, Fearnhead P, Hotter G, Mohan V, Collins-Emerson J, Kwan E, Besser TE, Cookson A, Carter PE and French NP, 2011. Whole-genome comparison of two *Campylobacter jejuni* isolates of the same sequence type reveals multiple loci of different ancestral lineage. PLoS One, 6, e27121.

Bosilevac JM and Koohmaraie M, 2011. Prevalence and characterization of non-O157 shiga toxin-producing *Escherichia coli* isolates from commercial ground beef in the United States. Appl Environ Microbiol, 77, 2103-2112.

Boxrud D, Pederson-Gulrud K, Wotton J, Medus C, Lyszkowicz E, Besser J and Bartkus JM, 2007. Comparison of multiple-locus variable-number tandem repeat analysis, pulsed-field gel electrophoresis, and phage typing for subtype analysis of *Salmonella enterica* serotype Enteritidis. J Clin Microbiol, 45, 536-543.

Boysen L, Rosenquist H, Larsson JT, Nielsen EM, Sorensen G, Nordentoft S and Hald T, 2013. Source attribution of human campylobacteriosis in Denmark. Epidemiol Infect, 1-10.

Bruun T, Soerensen G, Forshell LP, Jensen T, Nygard K, Kapperud G, Lindstedt BA, Wingstrand A, Petersen RF, Muller L, Kjelsoe C, Ivarsson S, Hjertqvist M, Lofdahl S and Ethelberg S, 2009. An outbreak of *Salmonella* Typhimurium infections in Denmark, Norway and Sweden, 2008. Eurosurveillance, 14 (1).

Buchholz U, Bernard H, Werber D, Bohmer MM, Remschmidt C, Wilking H, Delere Y, an der Heiden M, Adlhoch C, Dreesman J, Ehlers J, Ethelberg S, Faber M, Frank C, Fricke G, Greiner M, Hohle M, Ivarsson S, Jark U, Kirchner M, Koch J, Krause G, Luber P, Rosner B, Stark K and Kuhne M, 2011. German outbreak of *Escherichia coli* O104:H4 associated with sprouts. N Engl J Med, 365, 1763-1770.

Bugarel M, Vignaud ML, Moury F, Fach P and Brisabois A, 2012. Molecular identification in monophasic and nonmotile variants of *Salmonella enterica* serovar Typhimurium. Microbiologyopen, 1, 481-489.

Cao G, Meng J, Strain E, Stones R, Pettengill J, Zhao S, McDermott P, Brown E and Allard M, 2013. Phylogenetics and differentiation of *Salmonella* Newport lineages by whole genome sequencing. PLoS One, 8, e55687.

Carattoli A, 2013. Plasmids and the spread of resistance. Int J Med Microbiol, 303, 298-304.

Chakraborty T, Leimeister-Wachter M, Domann E, Hartl M, Goebel W, Nichterlein T and Notermans S, 1992. Coordinate regulation of virulence genes in *Listeria monocytogenes* requires the product of the prfA gene. J Bacteriol, 174, 568-574.

Chattopadhyay S, Taub F, Paul S, Weissman SJ and Sokurenko EV, 2013. Microbial variome database: point mutations, adaptive or not, in bacterial core genomes. Mol Biol Evol, 30, 1465-1470.

Chaudhuri RR, Morgan E, Peters SE, Pleasance SJ, Hudson DL, Davies HM, Wang J, van Diemen PM, Buckley AM, Bowen AJ, Pullinger GD, Turner DJ, Langridge GC, Turner AK, Parkhill J, Charles IG, Maskell DJ and Stevens MP, 2013. Comprehensive assignment of roles for *Salmonella* Typhimurium genes in intestinal colonization of food-producing animals. PLoS Genet, 9, e1003456.

Chen C, Zhang W, Zheng H, Lan R, Wang H, Du P, Bai X, Ji S, Meng Q, Jin D, Liu K, Jing H, Ye C, Gao GF, Wang L, Gottschalk M and Xu J, 2013. Minimum core genome sequence typing of bacterial pathogens: a unified approach for clinical and public health microbiology. J Clin Microbiol, 51, 2582-2591.

Chenal-Francisque V, Diancourt L, Cantinelli T, Passet V, Tran-Hykes C, Bracq-Dieye H, Leclercq A, Pourcel C, Lecuit M and Brisse S, 2013. Optimized Multilocus variable-number tandem-repeat analysis assay and its complementarity with pulsed-field gel electrophoresis and multilocus sequence typing for *Listeria monocytogenes* clone identification and surveillance. J Clin Microbiol, 51, 1868-1880.

Cheung MK and Kwan HS, 2012. Fighting outbreaks with bacterial genomics: case review and workflow proposal. Public Health Genomics, 15, 341-351.

Cho S, Whittam TS, Boxrud DJ, Bartkus JM, Rankin SC, Wilkins MJ, Somsel P, Downes FP, Musser KA, Root TP, Warnick LD, Wiedmann M and Saeed AM, 2010. Use of multiple-locus variable number tandem repeat analysis and phage typing for subtyping of *Salmonella* Enteritidis from sporadic human cases in the United States. J Appl Microbiol, 108, 859-867.

Clark CG, Kruk TM, Bryden L, Hirvi Y, Ahmed R and Rodgers FG, 2003. Subtyping of *Salmonella enterica* serotype Enteritidis strains by manual and automated PstI-SphI ribotyping. J Clin Microbiol, 41, 27-33.

Clayton EM, Hill C, Cotter PD and Ross RP, 2011. Real-Time PCR Assay To Differentiate Listeriolysin S-Positive and -Negative Strains of *Listeria monocytogenes*.

Cody AJ, McCarthy ND, Jansen van Rensburg M, Isinkaye T, Bentley SD, Parkhill J, Dingle KE, Bowler IC, Jolley KA and Maiden MC, 2013. Real-time genomic epidemiological evaluation of human *Campylobacter* isolates by use of whole-genome multilocus sequence typing. J Clin Microbiol, 51, 2526-2534.

Colles FM and Maiden MC, 2012. *Campylobacter* sequence typing databases: applications and future prospects. Microbiology, 158, 2695-2709.

Craun GF, Brunkard JM, Yoder JS, Roberts VA, Carpenter J, Wade T, Calderon RL, Roberts JM, Beach MJ and Roy SL, 2010. Causes of outbreaks associated with drinking water in the United States from 1971 to 2006. Clin Microbiol Rev, 23, 507-528.

Crayford G, Wigley P and Humphrey T, 2011. The infection biology of pig-associated *Salmonella*. 9th International Conference on the Epidemiology and Control of biological, chemical and physical hazards in pigs and pork (SafePork 2011), Maastricht, The Netherlands, 19-22 June 2011. 57-61.

Crerar SK, Castle M, Hassel S and Schumacher D, 2011. Recent Experiences with *Listeria monocytogenes* in New Zealand and development of a food control risk-based strategy. Food Control, 22(9), 1510-1512.

Dasti JI, Tareen AM, Lugert R, Zautner AE and Gross U, 2010. *Campylobacter jejuni*: a brief overview on pathogenicity-associated factors and disease-mediating mechanisms. Int J Med Microbiol, 300, 205-211.

David JM, Guillemot D, Bemrah N, Thebault A, Brisabois A, Chemaly M, Weill FX, Sanders P and Watier L, 2012. The Bayesian microbial subtyping attribution model: robustness to prior information and a proposition. Risk Anal, 33, 397-408.

David JM, Sanders P, Bemrah N, Granier SA, Denis M, Weill FX, Guillemot D and Watier L, 2013. Attribution of the French human Salmonellosis cases to the main food-sources according to the type of surveillance data. Prev Vet Med, 110, 12-27.

Davies R, Deuchande R, Larki L, Collins R and Irvine RM, 2013. Multidrug resistant *Salmonella* Java found in British broiler flocks. Vet Rec, 172, 617-618.

DebRoy C, Roberts E and Fratamico PM, 2011. Detection of O antigens in *Escherichia coli*. Anim Health Res Rev, 12, 169-185.

den Bakker HC, Cummings CA, Ferreira V, Vatta P, Orsi RH, Degoricija L, Barker M, Petrauskene O, Furtado MR and Wiedmann M, 2010. Comparative genomics of the bacterial genus *Listeria*: Genome evolution is characterized by limited gene acquisition and limited gene loss. BMC Genomics, 11, 688.

Denny J, Threlfall J, Takkinen J, Lofdahl S, Westrell T, Varela C, Adak B, Boxall N, Ethelberg S, Torpdahl M, Straetemans M and van Pelt W, 2007. Multinational *Salmonella* Paratyphi B variant Java (*Salmonella* Java) outbreak, August - December 2007. Euro Surveill, 12, E071220 071222.

Desai PT, Porwollik S, Long F, Cheng P, Wollam A, Bhonagiri-Palsikar V, Hallsworth-Pepin K, Clifton SW, Weinstock GM and McClelland M, 2013. Evolutionary Genomics of *Salmonella enterica* Subspecies. MBio, 4.

Dewaele I, Rasschaert G, Bertrand S, Wildemauwe C, Wattiau P, Imberechts H, Herman L, Ducatelle R, De Reu K and Heyndrickx M, 2012. Molecular characterization of *Salmonella* Enteritidis: comparison of an optimized multi-locus variable-number of tandem repeat analysis (MLVA) and pulsed-field gel electrophoresis. Foodborne Pathog Dis, 9, 885-895.

Dhar MS and Virdi JS, 2013. Strategies used by *Yersinia enterocolitica* to evade killing by the host: Thinking beyond Yops. Microbes Infect.

Diaz-Sanchez S, Hanning I, Pendleton S and D'Souza D, 2013. Next-generation sequencing: the future of molecular genetics in poultry production and food safety. Poult Sci, 92, 562-572.

Didelot X, Bowden R, Street T, Golubchik T, Spencer C, McVean G, Sangal V, Anjum MF, Achtman M, Falush D and Donnelly P, 2011. Recombination and population structure in *Salmonella enterica*. PLoS Genet, 7, e1002191.

Didelot X and Maiden MC, 2010. Impact of recombination on bacterial evolution. Trends Microbiol, 18, 315-322.

Dimarzio M, Shariat N, Kariyawasam S, Barrangou R and Dudley EG, 2013. Antibiotic resistance in *Salmonella* Typhimurium associates with CRISPR sequence type. Antimicrob Agents Chemother.

Dingle KE, Colles FM, Ure R, Wagenaar JA, Duim B, Bolton FJ, Fox AJ, Wareing DR and Maiden MC, 2002. Molecular characterization of *Campylobacter jejuni* clones: a basis for epidemiologic investigation. Emerg Infect Dis, 8, 949-955.

Dingle KE, Colles FM, Wareing DR, Ure R, Fox AJ, Bolton FE, Bootsma HJ, Willems RJ, Urwin R and Maiden MC, 2001. Multilocus sequence typing system for *Campylobacter jejuni*. J Clin Microbiol, 39, 14-23.

Dobrindt U, Hochhut B, Hentschel U and Hacker J, 2004. Genomic islands in pathogenic and environmental microorganisms. Nat Rev Microbiol, 2, 414-424.

Doumith M, Buchrieser C, Glaser P, Jacquet C and Martin P, 2004. Differentiation of the major *Listeria monocytogenes* serovars by multiplex PCR. J Clin Microbiol, 42, 3819-3822.

Doumith M, Jacquet C, Gerner-Smidt P, Graves LM, Loncarevic S, Mathisen T, Morvan A, Salcedo C, Torpdahl M, Vazquez JA and Martin P, 2005. Multicenter validation of a multiplex PCR assay for differentiating the major *Listeria monocytogenes* serovars 1/2a, 1/2b, 1/2c, and 4b: toward an international standard. J Food Prot, 68, 2648-2650.

ECDC (European Centre for Disease Prevention and Control), 2007. Surveillance of communicable diseases in the European Union. A long-term strategy | 2008–2013. (2007) Available online: http://www.ecdc.europa.eu/en/aboutus/key%20documents/08-13_kd_surveillance_of_cd.pdf.

ECDC (European Centre for Disease Prevention and Control), 2013. Surveillance of communicable diseases in Europe - a concept to integrate molecular typing data into EU-level surveillance. Version 2.4. 7 September 2011. Available online at: http://www.ecdc.europa.eu/en/publications/Publications/surveillance-concept-molecular%20typing-sept2011.pdf (last accessed on 13/12/2013).

EFSA (European Food Safety Authority), 2008. Scientific Opinion of the Panel on Biological Hazards on overview of methods for source attribution for human illness from food-borne microbiological hazards. The EFSA Journal 2008, 764, 1-43.

EFSA (European Food Safety Authority), 2009. Report on the availability of molecular typing methods for *Salmonella*, *Campylobacter*, *verotoxigenic Escherichia coli*, *Listeria monocytogenes* and *Staphylococcus aureus* isolates from food, animals and feedingstuffs in European Union Member States (and in some other reporting countries). The EFSA Journal 2009, 272r, 1-52.

EFSA Panel on Biological Hazards (BIOHAZ), 2010. Scientific Opinion on monitoring and assessment of the public health risk of '*Salmonella* Typhimurium-like' strains. EFSA Journal 2010;8(10):1826, 48 pp. doi:10.2903/j.efsa.2010.1826

EFSA Panel on Biological Hazards (BIOHAZ), 2011a. Scientific Opinion on *Campylobacter* in broiler meat production: control options and performance objectives and/or targets at different stages of the food chain. EFSA Journal 2011;9(4):2105, 141 pp. doi:10.2903/j.efsa.2011.2105

EFSA Panel on Biological Hazards (BIOHAZ), 2011b. Scientific Opinion on a quantitative estimation of the public health impact of setting a new target for the reduction of *Salmonella* in broilers. EFSA Journal 2011;9(7):2016, 94 pp. doi:10.2903/j.efsa.2011.2106

EFSA Panel on Biological Hazards (BIOHAZ), 2012. Scientific Opinion on an estimation of the public health impact of setting a new target for the reduction of *Salmonella* in turkeys. EFSA Journal 2012;10(4):2616, 89 pp. doi:10.2903/j.efsa.2012.2616

EFSA BIOHAZ Panel (EFSA Panel on Biological Hazards), 2013. Scientific Opinion on VTEC-seropathotype and scientific criteria regarding pathogenicity assessment. EFSA Journal 2013;11(4):3138, 106 pp. doi:10.2903/j.efsa.2013.3138

EFSA and ECDC (European Food Safety Authority and European Centre for Disease Prevention and Control), 2012. The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in 2010. EFSA Journal 2012;10(3):2597, 442 pp. doi:10.2903/j.efsa.2012.2597

EFSA and ECDC (European Food Safety Authority and European Centre for Disease Prevention and Control), 2013. The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in 2011. EFSA Journal 2013;11(4):3129, 250 pp. doi:10.2903/j.efsa.2013.3129

Ellström P, Feodoroff B, Hanninen ML and Rautelin H, 2013. Characterization of clinical *Campylobacter jejuni* isolates with special emphasis on lipooligosaccharide locus class, putative virulence factors and host response. Int J Med Microbiol, 303, 134-139.

Fabre L, Zhang J, Guigon G, Le Hello S, Guibert V, Accou-Demartin M, de Romans S, Lim C, Roux C, Passet V, Diancourt L, Guibourdenche M, Issenhuth-Jeanjean S, Achtman M, Brisse S, Sola C and Weill FX, 2012. CRISPR typing and subtyping for improved laboratory surveillance of *Salmonella* infections. PLoS One, 7, e36995.

Fabrega A and Vila J, 2013. *Salmonella enterica* serovar Typhimurium skills to succeed in the host: virulence and regulation. Clin Microbiol Rev, 26, 308-341.

Feil EJ, Li BC, Aanensen DM, Hanage WP and Spratt BG, 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol, 186, 1518-1530.

Felix B, Dao TT, Grout J, Lombard B, Assere A, Brisabois A and Roussel S, 2012. Pulsed-field gel electrophoresis, conventional, and molecular serotyping of *Listeria monocytogenes* from food proficiency testing trials toward an harmonization of subtyping at European level. Foodborne Pathog Dis, 9, 719-726.

Fischer J, Rodriguez I, Schmoger S, Friese A, Roesler U, Helmuth R and Guerra B, 2012. *Escherichia coli* producing VIM-1 carbapenemase isolated on a pig farm. The Journal of antimicrobial chemotherapy, 67, 1793-1795.

Fischer J, Rodriguez I, Schmoger S, Friese A, Roesler U, Helmuth R and Guerra B, 2013a. *Salmonella enterica* subsp *enterica* producing VIM-1 carbapenemase isolated from livestock farms. Journal of Antimicrobial Chemotherapy, 68, 478-480.

Fischer J, Schmoger S, Baumann B, Laube H, von Salviati C, Friese A, Roesler U, Helmuth R and B. G, 2013b. Carbapenemases are present in enterobacteria from non-domestic animals...more frequently than we thought. In: Abstracts of the 23rd European Congress of Clinical Microbiology and Infectious Disease, ECCMID, Berlin, Germany. P1449. European Society for Clinical Microbiology and Infectious Disease, Basel, Switzerland.

Fisher IS and Threlfall EJ, 2005. The Enter-net and Salm-gene databases of foodborne bacterial pathogens that cause human infections in Europe and beyond: an international collaboration in surveillance and the development of intervention strategies. Epidemiol Infect, 133, 1-7.

Fitzgerald C, Collins M, van Duyne S, Mikoleit M, Brown T and Fields P, 2007. Multiplex, bead-based suspension array for molecular determination of common *Salmonella* serogroups. J Clin Microbiol, 45, 3323-3334.

Fitzgerald C, Helsel LO, Nicholson MA, Olsen SJ, Swerdlow DL, Flahart R, Sexton J and Fields PI, 2001. Evaluation of methods for subtyping *Campylobacter jejuni* during an outbreak involving a food handler. J Clin Microbiol, 39, 2386-2390.

Foley SL, Lynne AM and Nayak R, 2009. Molecular typing methodologies for microbial source tracking and epidemiological investigations of Gram-negative bacterial foodborne pathogens. Infect Genet Evol, 9, 430-440.

Frank C, Faber MS, Askar M, Bernard H, Fruth A, Gilsdorf A, Hohle M, Karch H, Krause G, Prager R, Spode A, Stark K and Werber D, 2011a. Large and ongoing outbreak of haemolytic uraemic syndrome, Germany, May 2011. Euro Surveill, 16.

Frank C, Werber D, Cramer JP, Askar M, Faber M, an der Heiden M, Bernard H, Fruth A, Prager R, Spode A, Wadl M, Zoufaly A, Jordan S, Kemper MJ, Follin P, Muller L, King LA, Rosner B, Buchholz U, Stark K and Krause G, 2011b. Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104:H4 outbreak in Germany. N Engl J Med, 365, 1771-1780.

Fratamico PM, Bagi LK, Cray WC, Jr., Narang N, Yan X, Medina M and Liu Y, 2011. Detection by multiplex real-time polymerase chain reaction assays and isolation of Shiga toxin-producing *Escherichia coli* serogroups O26, O45, O103, O111, O121, and O145 in ground beef. Foodborne Pathog Dis, 8, 601-607.

Friedrich AW, Bielaszewska M, Zhang WL, Pulz M, Kuczius T, Ammon A and Karch H, 2002. *Escherichia coli* harboring Shiga toxin 2 gene variants: frequency and association with clinical symptoms. J Infect Dis, 185, 74-84.

Garcia-Fernandez A, Chiaretto G, Bertini A, Villa L, Fortini D, Ricci A and Carattoli A, 2008. Multilocus sequence typing of IncI1 plasmids carrying extended-spectrum beta-lactamases in *Escherichia coli* and *Salmonella* of human and animal origin. J Antimicrob Chemother, 61, 1229-1233.

Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, Brodkin E, Rempel S, Moore R, Zhao Y, Holt R, Varhol R, Birol I, Lem M, Sharma MK, Elwood K, Jones SJ, Brinkman FS, Brunham RC and Tang P, 2011. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N Engl J Med, 364, 730-739.

Gebreyes W, 2011. Use of heavy metals in swine feed and its association with the co-selection of metal tolerant and multi-drug resistant *Salmonella*. 9th International Conference on the Epidemiology and Control of biological, chemical and physical hazards in pigs and pork (SafePork 2011), Maastricht, the Netherlands, 19-22 June 2011. 186-188.

Gerner-Smidt P, Kincaid J, Kubota K, Hise K, Hunter SB, Fair MA, Norton D, Woo-Ming A, Kurzynski T, Sotir MJ, Head M, Holt K and Swaminathan B, 2005. Molecular surveillance of shiga toxigenic *Escherichia coli* O157 by PulseNet USA. J Food Prot, 68, 1926-1931.

Gerner-Smidt P and Scheutz F, 2006. Standardized pulsed-field gel electrophoresis of Shiga toxin-producing *Escherichia coli*: the PulseNet Europe Feasibility Study. Foodborne Pathog Dis, 3, 74-80.

Graves LM, Hunter SB, Ong AR, Schoonmaker-Bopp D, Hise K, Kornstein L, DeWitt WE, Hayes PS, Dunne E, Mead P and Swaminathan B, 2005. Microbiological aspects of the investigation that traced the 1998 outbreak of listeriosis in the United States to contaminated hot dogs and establishment of molecular subtyping-based surveillance for *Listeria monocytogenes* in the PulseNet network. J Clin Microbiol, 43, 2350-2355.

Graves LM and Swaminathan B, 2001. PulseNet standardized protocol for subtyping *Listeria monocytogenes* by macrorestriction and pulsed-field gel electrophoresis. Int J Food Microbiol, 65, 55-62.

Griekspoor P, Colles FM, McCarthy ND, Hansbro PM, Ashhurst-Smith C, Olsen B, Hasselquist D, Maiden MC and Waldenstrom J, 2013. Marked host specificity and lack of phylogeographic population structure of *Campylobacter jejuni* in wild birds. Mol Ecol, 22, 1463-1472.

Grif K, Heller I, Wagner M, Dierich M and Wurzner R, 2006. A comparison of *Listeria monocytogenes* serovar 4b isolates of clinical and food origin in Austria by automated ribotyping and pulsed-field gel electrophoresis. Foodborne Pathog Dis, 3, 138-141.

Guerra B, Soto SM, Arguelles JM and Mendoza MC, 2001. Multidrug resistance is mediated by large plasmids carrying a class 1 integron in the emergent *Salmonella enterica* serotype [4,5,12:i:-]. Antimicrob Agents Chemother, 45, 1305-1308.

Guiney DG and Fierer J, 2011. The Role of the spv Genes in *Salmonella* Pathogenesis. Front Microbiol, 2, 129.

Guo C, Hoekstra RM, Schroeder CM, Pires SM, Ong KL, Hartnett E, Naugle A, Harman J, Bennett P, Cieslak P, Scallan E, Rose B, Holt KG, Kissler B, Mbandi E, Roodsari R, Angulo FJ and Cole D, 2011. Application of Bayesian techniques to model the burden of human salmonellosis attributable to U.S. food commodities at the point of processing: adaptation of a Danish model. Foodborne Pathog Dis, 8, 509-516.

Guo D, Liu B, Liu F, Cao B, Chen M, Hao X, Feng L and Wang L, 2013. Development of a DNA microarray for molecular identification of all 46 *Salmonella* O serogroups. Appl Environ Microbiol, 79, 3392-3399.

Gupta S and Maiden MCJ, 2001. Exploring the evolution of diversity in pathogen populations. Trends Microbiol 9, 181-192.

Hacker J and Carniel E, 2001. Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. EMBO Rep, 2, 376-381.

Hald T, Lo Fo Wong DM and Aarestrup FM, 2007. The attribution of human infections with antimicrobial resistant *Salmonella* bacteria in Denmark to sources of animal origin. Foodborne Pathog Dis, 4, 313-326.

Hald T and Lund J, 2012. Development of a user-friendly interface version of the *Salmonella* source-attribution model. External Scientific Report. Supporting Publications 2012:EN-318. [77 pp.].

Hald T, Vose D, Wegener HC and Koupeev T, 2004. A Bayesian approach to quantify the contribution of animal-food sources to human salmonellosis. Risk Anal, 24, 255-269.

Hall RM, 2010. *Salmonella* genomic islands and antibiotic resistance in *Salmonella enterica*. Future Microbiol, 5, 1525-1538.

Hallin M, Deplano A and Struelens MJ 2012. Molecular typing of bacterial pathogens: a tool for the epidemiological study and control of infectious diseases. In: New Frotiers of Molecular Epidemiology of Infectious diseases. Editors: Morand, s., Beaudeau, F. and Cabaret, J. Springer, London. ISBN 978-94-007-2114-2

Hammarlöf DL, Canals R and Hinton JC, 2013. The FUN of identifying gene function in bacterial pathogens; insights from *Salmonella* functional genomics. Curr Opin Microbiol.

Hannemann S, Gao B and Galan JE, 2013. *Salmonella* modulation of host cell gene expression promotes its intracellular growth. PLoS Pathog, 9, e1003668.

Haque F, Li J, Wu HC, Liang XJ and Guo P, 2013. Solid-State and Biological Nanopore for Real-Time Sensing of Single Chemical and Sequencing of DNA. Nano Today, 8, 56-74.

Harrington CS, Thomson-Carter FM and Carter PE, 1997. Evidence for recombination in the flagellin locus of *Campylobacter jejuni*: implications for the flagellin gene typing scheme. J Clin Microbiol, 35, 2386-2392.

Hauser E, Mellmann A, Semmler T, Stoeber H, Wieler LH, Karch H, Kuebler N, Fruth A, Harmsen D, Weniger T, Tietze E and Schmidt H, 2013. Phylogenetic and molecular analysis of food-borne shiga toxin-producing *Escherichia coli*. Appl Environ Microbiol, 79, 2731-2740.

Havelaar AH, Braunig J, Christiansen K, Cornu M, Hald T, Mangen MJ, Molbak K, Pielaat A, Snary E, Van Pelt W, Velthuis A and Wahlstrom H, 2007. Towards an integrated approach in supporting microbiological food safety decisions. Zoonoses Public Health, 54, 103-117.

Hazen TH, Sahl JW, Fraser CM, Donnenberg MS, Scheutz F and Rasko DA, 2013. Refining the pathovar paradigm via phylogenomics of the attaching and effacing *Escherichia coli*. Proc Natl Acad Sci U S A, 110, 12810-12815.

Hopkins KL, Kirchner M, Guerra B, Granier SA, Lucarelli C, Porrero MC, Jakubczak A, Threlfall EJ and Mevius DJ, 2010. Multiresistant *Salmonella enterica* serovar 4,[5], 12:i:- in Europe: a new pandemic strain? Eurosurveillance, 15, 2-10.

Hopkins KL, Liebana E, Villa L, Batchelor M, Threlfall EJ and Carattoli A, 2006. Replicon typing of plasmids carrying CTX-M or CMY beta-lactamases circulating among *Salmonella* and *Escherichia coli* isolates. Antimicrob Agents Chemother, 50, 3203-3206.

Hopkins KL, Peters TM, de Pinna E and Wain J, 2011. Standardisation of multilocus variable-number tandem-repeat analysis (MLVA) for subtyping of *Salmonella enterica* serovar Enteritidis. Euro Surveill, 16.

Humphrey B, Thomson NR, Thomas CM, Brooks K, Sanders M, Delsol AA, Roe JM, Bennett PM and Enne VI, 2012. Fitness of *Escherichia coli* strains carrying expressed and partially silent IncN and IncP1 plasmids. BMC Microbiol, 12, 53.

Hunter PR, 1990. Reproducibility and indices of discriminatory power of microbial typing methods. J Clin Microbiol, 28, 1903-1905.

Hunter PR and Gaston MA, 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol, 26, 2465-2466.

Hyytia-Trees E, Smole SC, Fields PA, Swaminathan B and Ribot EM, 2006. Second generation subtyping: a proposed PulseNet protocol for multiple-locus variable-number tandem repeat analysis of Shiga toxin-producing *Escherichia coli* O157 (STEC O157). Foodborne Pathog Dis, 3, 118-131.

Ibarra JA and Steele-Mortimer O, 2009. *Salmonella*-the ultimate insider. *Salmonella* virulence factors that modulate intracellular survival. Cell Microbiol, 11, 1579-1586.

ICMSF 1996. Microorganisms in foods 5. Microbiological specifications for of food pathogens: *Listeria monocytogenes*. p. 141-182. Blackie Academic & Professional, UK.

Ji XW, Liao YL, Zhu YF, Wang HG, Gu L, Gu J, Dong C, Ding HL, Mao XH, Zhu FC and Zou QM, 2010. Multilocus sequence typing and virulence factors analysis of *Escherichia coli* O157 strains in China. J Microbiol, 48, 849-855.

Johnson TJ, Bielak EM, Fortini D, Hansen LH, Hasman H, Debroy C, Nolan LK and Carattoli A, 2012. Expansion of the IncX plasmid family for improved identification and typing of novel plasmids in drug-resistant Enterobacteriaceae. Plasmid, 68, 43-50.

Kaper JB, Nataro JP and Mobley HL, 2004. Pathogenic *Escherichia coli*. Nat Rev Microbiol, 2, 123-140.

Karch H, Denamur E, Dobrindt U, Finlay BB, Hengge R, Johannes L, Ron EZ, Tonjum T, Sansonetti PJ and Vicente M, 2012. The enemy within us: lessons from the 2011 European *Escherichia coli* O104:H4 outbreak. EMBO Mol Med, 4, 841-848.

Karmali MA, 2003. The medical significance of Shiga toxin-producing *Escherichia coli* infections. An overview. Methods Mol Med, 73, 1-7.

Kerouanton A, Marault M, Petit L, Grout J, Dao TT and Brisabois A, 2010. Evaluation of a multiplex PCR assay as an alternative method for *Listeria monocytogenes* serotyping. J Microbiol Methods, 80, 134-137.

Knabel SJ, Reimer A, Verghese B, Lok M, Ziegler J, Farber J, Pagotto F, Graham M, Nadon CA and Gilmour MW, 2012. Sequence typing confirms that a predominant *Listeria monocytogenes* clone caused human listeriosis cases and outbreaks in Canada from 1988 to 2010. J Clin Microbiol, 50, 1748-1751.

Krishnaswami B, Jandhyala S, Lou Y, Siddiqui A, Agate M, Dhapulkar A, Breu H, Shukla A, Kuhlmann K, Hyland F, Altun G and DThomas D, 2013. Systems and methods for analysis and interpretation of nucleic acid sequence data. U.S. Patent 20,130,091,126, issued April 11, 2013. Available online at: http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=/netahtml/PTO/srchnum.html&r=1&f=G&l=50&s1=20130091126.PGNR.

Lan R, Reeves PR and Octavia S, 2009. Population structure, origins and evolution of major *Salmonella enterica* clones. Infect Genet Evol, 9, 996-1005.

Larsson JT, Torpdahl M and Moller Nielsen E, 2013. Proof-of-concept study for successful inter-laboratory comparison of MLVA results. Euro Surveill, 18, 20566.

Larsson JT, Torpdahl M, Petersen RF, Sorensen G, Lindstedt BA and Nielsen EM, 2009. Development of a new nomenclature for *Salmonella* Typhimurium multilocus variable number of tandem repeats analysis (MLVA). Euro Surveill, 14.

Lawson AJ, Desai M, O'Brien SJ, Davies RH, Ward LR and Threlfall EJ, 2004. Molecular characterisation of an outbreak strain of multiresistant *Salmonella enterica* serovar Typhimurium DT104 in the UK. Clin Microbiol Infect, 10, 143-147.

Leekitcharoenphon P, Lukjancenko O, Friis C, Aarestrup FM and Ussery DW, 2012. Genomic variation in *Salmonella enterica* core genes for epidemiological typing. BMC Genomics, 13, 88.

Li W, Raoult D and Fournier PE, 2009. Bacterial strain typing in the genomic era. FEMS Microbiol Rev, 33, 892-916.

Liebana E, Clouting C, Garcia-Migura L, Clifton-Hadley FA, Lindsay E, Threlfall EJ and Davies RH, 2004. Multiple genetic typing of *Salmonella* Enteritidis phage-types 4, 6, 7, 8 and 13a isolates from animals and humans in the UK. Vet Microbiol, 100, 189-195.

Liebana E, Garcia-Migura L, Clouting C, Clifton-Hadley FA, Lindsay E, Threlfall EJ, McDowell SW and Davies RH, 2002. Multiple genetic typing of *Salmonella enterica* serotype Typhimurium isolates of different phage types (DT104, U302, DT204b, and DT49) from animals and humans in England, Wales, and Northern Ireland. J Clin Microbiol, 40, 4450-4456.

Lindstedt BA, Tham W, Danielsson-Tham ML, Vardund T, Helmersson S and Kapperud G, 2008. Multiple-locus variable-number tandem-repeats analysis of *Listeria monocytogenes* using multicolour capillary electrophoresis and comparison with pulsed-field gel electrophoresis typing. J Microbiol Methods, 72, 141-148.

Lindstedt BA, Torpdahl M, Vergnaud G, Le Hello S, Weill FX, Tietze E, Malorny B, Prendergast DM, Ni Ghallchoir E, Lista RF, Schouls LM, Soderlund R, Borjesson S and Akerstrom S, 2013. Use of multilocus variable-number tandem repeat analysis (MLVA) in eight European countries, 2012. Euro Surveill, 18, 20385.

Lindstedt BA, Vardund T, Aas L and Kapperud G, 2004a. Multiple-locus variable-number tandem-repeats analysis of *Salmonella enterica* subsp. *enterica* serovar Typhimurium using PCR multiplexing and multicolor capillary electrophoresis. J Microbiol Methods, 59, 163-172.

Lindstedt BA, Vardund T and Kapperud G, 2004b. Multiple-Locus Variable-Number Tandem-Repeats Analysis of *Escherichia coli* O157 using PCR multiplexing and multi-colored capillary electrophoresis. J Microbiol Methods, 58, 213-222.

Little CL, Pires SM, Gillespie IA, Grant K and Nichols GL, 2010. Attribution of human *Listeria monocytogenes* infections in England and Wales to ready-to-eat food sources placed on the market: adaptation of the Hald *Salmonella* source attribution model. Foodborne Pathog Dis, 7, 749-756.

Liu F, Barrangou R, Gerner-Smidt P, Ribot EM, Knabel SJ and Dudley EG, 2011a. Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme for subtyping of the major serovars of *Salmonella enterica* subsp. *enterica*. Appl Environ Microbiol, 77, 1946-1956.

Liu F, Kariyawasam S, Jayarao BM, Barrangou R, Gerner-Smidt P, Ribot EM, Knabel SJ and Dudley EG, 2011b. Subtyping *Salmonella enterica* serovar Enteritidis isolates from different sources by using sequence typing based on virulence genes and clustered regularly interspaced short palindromic repeats (CRISPRs). Appl Environ Microbiol, 77, 4520-4526.

Lobel L, Sigal N, Borovok I, Ruppin E and Herskovits AA, 2012. Integrative genomic analysis identifies isoleucine and CodY as regulators of *Listeria monocytogenes* virulence. PLoS Genet, 8, e1002887.

Lomonaco S, Patti R, Knabel SJ and Civera T, 2012. Detection of virulence-associated genes and epidemic clone markers in *Listeria monocytogenes* isolates from PDO Gorgonzola cheese. Int J Food Microbiol, 160, 76-79.

Lorenz MG and Wackernagel W, 1994. Bacterial gene transfer by natural genetic transformation in the environment. Microbiol Rev, 58, 563-602.

Ma H and Bryers JD, 2013. Non-invasive determination of conjugative transfer of plasmids bearing antibiotic-resistance genes in biofilm-bound bacteria: effects of substrate loading and antibiotic selection. Appl Microbiol Biotechnol, 97, 317-328.

MacCannell D, 2013. Bacterial strain typing. Clin Lab Med, 33, 629-650.

Magnusson SH, Guethmundsdottir S, Reynisson E, Runarsson AR, Harethardottir H, Gunnarson E, Georgsson F, Reiersen J and Marteinsson VT, 2011. Comparison of *Campylobacter jejuni* isolates from human, food, veterinary and environmental sources in Iceland using PFGE, MLST and fla-SVR sequencing. J Appl Microbiol, 111, 971-981.

Maiden MC, 2006. Multilocus sequence typing of bacteria. Annu Rev Microbiol, 60, 561-588.

Maiden MC, van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA and McCarthy ND, 2013. MLST revisited: the gene-by-gene approach to bacterial genomics. Nat Rev Microbiol, 11, 728-736.

Manfreda G, De Cesare A, Bondioli V and Franchini A, 2003. Ribotyping characterisation of *Campylobacter* isolates randomly collected from different sources in Italy. Diagn Microbiol Infect Dis, 47, 385-392.

Marotta F, Zilli K, Tonelli A, Sacchini L, Alessiani A, Migliorati G and Di Giannatale E, 2013. Detection and genotyping of *Campylobacter jejuni* and *Campylobacter coli* by use of DNA oligonucleotide arrays. Mol Biotechnol, 53, 182-188.

Martin IE, Tyler SD, Tyler KD, Khakhria R and Johnson WM, 1996. Evaluation of ribotyping as epidemiologic tool for typing *Escherichia coli* serogroup O157 isolates. J Clin Microbiol, 34, 720-723.

Martin P, Jacquet C, Goulet V, Vaillant V and De Valk H, 2006. Pulsed-field gel electrophoresis of *Listeria monocytogenes* strains: the PulseNet Europe Feasibility Study. Foodborne Pathog Dis, 3, 303-308.

Mather AE, Reid SW, Maskell DJ, Parkhill J, Fookes MC, Harris SR, Brown DJ, Coia JE, Mulvey MR, Gilmour MW, Petrovska L, de Pinna E, Kuroda M, Akiba M, Izumiya H, Connor TR, Suchard MA, Lemey P, Mellor DJ, Haydon DT and Thomson NR, 2013. Distinguishable epidemics of multidrug-resistant *Salmonella* Typhimurium DT104 in different hosts. Science, 341, 1514-1517.

Maynard Smith J, Simth NH, O'Rourke M and Spratt BG, 1993. How clonal are bacteria? Proc Natl Acad Sci USA 90, 4384-4388.

McCarthy ND, Colles FM, Dingle KE, Bagnall MC, Manning G, Maiden MC and Falush D, 2007. Host-associated genetic import in *Campylobacter jejuni*. Emerg Infect Dis, 13, 267-272.

McQuiston JR, Waters RJ, Dinsmore BA, Mikoleit ML and Fields PI, 2011. Molecular determination of H antigens of *Salmonella* by use of a microsphere-based liquid array. J Clin Microbiol, 49, 565-573.

Meinersmann RJ, Helsel LO, Fields PI and Hiett KL, 1997. Discrimination of *Campylobacter jejuni* isolates by fla gene sequencing. J Clin Microbiol, 35, 2810-2814.

Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, Prior K, Szczepanowski R, Ji Y, Zhang W, McLaughlin SF, Henkhaus JK, Leopold B, Bielaszewska M, Prager R, Brzoska PM, Moore RL, Guenther S, Rothberg JM and Karch H, 2011. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. PLoS One, 6, e22751.

Melo RT, Nalevaiko PC, Mendonça EP, Borges LW, Fonseca BB and Beletti M, 2013. *Campylobacter jejuni* strains isolated from chicken meat harbour several virulence factors and represent a potential risk to humans. Food Control, 33(1), 227-231.

Miko A, Lindstedt BA, Brandal LT, Lobersli I and Beutin L, 2010. Evaluation of multiple-locus variable number of tandem-repeats analysis (MLVA) as a method for identification of clonal groups among enteropathogenic, enterohaemorrhagic and avirulent *Escherichia coli* O26 strains. FEMS Microbiol Lett, 303, 137-146.

Miller JM, 2013. Whole-genome mapping: a new paradigm in strain-typing technology. J Clin Microbiol, 51, 1066-1070.

Monk JM, Charusani P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, Fesit AM and Palsson BØ, 2013. Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. Proceedings of the National Academy of Sciences of the United States of America. Published online before print November 25, 2013, doi: 10.1073/pnas.1307797110

Mossong J, Ragimbeau C, Schuh J, Weicherding P, Peetso R, Wildemauwe C, Imberechts H, Rabsch W and Bertrand S, 2012. Investigation of an excess of *Salmonella* Enteritidis phage type 14b and MLVA type 4-7-3-13-10-2-2 in Luxembourg, Belgium and Germany during 2010. Bull Soc Sci Med Grand Duche Luxemb, 49-62.

Moxon ER, Rainey PB, Nowak MA and Lenski RE, 1994. Adaptive evolution of highly mutable loci in pathogenic bacteria. Curr Biol, 4, 24-33.

Muellner P, Pleydell E, Pirie R, Baker MG, Campbell D, Carter PE and French NP, 2013. Molecular-based surveillance of campylobacteriosis in New Zealand--from source attribution to genomic epidemiology. Euro Surveill, 18.

Mughini Gras L, Smid JH, Wagenaar JA, de Boer AG, Havelaar AH, Friesema IH, French NP, Busani L and van Pelt W, 2012. Risk factors for campylobacteriosis of chicken, ruminant, and environmental origin: a combined case-control and source attribution analysis. PLoS One, 7, e42599.

Müller A, Rychli K, Muhterem-Uyar M, Zaiser A, Stessl B, Guinane CM, Cotter PD, Wagner M and Schmitz-Esser S, 2013. Tn6188 - A Novel Transposon in *Listeria monocytogenes* Responsible for Tolerance to Benzalkonium Chloride. PLoS One, 8, e76835.

Mullner P, Jones G, Noble A, Spencer SE, Hathaway S and French NP, 2009. Source attribution of food-borne zoonoses in New Zealand: a modified Hald model. Risk Anal, 29, 970-984.

Nadon CA, Trees E, Ng LK, Moller Nielsen E, Reimer A, Maxwell N, Kubota KA and Gerner-Smidt P, 2013. Development and application of MLVA methods as a tool for inter-laboratory surveillance. Euro Surveill, 18, 20565.

Nataro JP and Kaper JB, 1998. Diarrheagenic *Escherichia coli*. Clin Microbiol Rev, 11, 142-201.

Nataro JP, Steiner T and Guerrant RL, 1998. Enteroaggregative *Escherichia coli*. Emerg Infect Dis, 4, 251-261.

Nelson KE, Fouts DE, Mongodin EF, Ravel J, DeBoy RT, Kolonay JF, Rasko DA, Angiuoli SV, Gill SR, Paulsen IT, Peterson J, White O, Nelson WC, Nierman W, Beanan MJ, Brinkac LM, Daugherty SC, Dodson RJ, Durkin AS, Madupu R, Haft DH, Selengut J, Van Aken S, Khouri H, Fedorova N, Forberger H, Tran B, Kathariou S, Wonderling LD, Uhlich GA, Bayles DO, Luchansky JB and Fraser CM, 2004. Whole genome comparisons of serotype 4b and 1/2a strains of the food-borne pathogen *Listeria monocytogenes* reveal new insights into the core genome components of this species. Nucleic Acids Res, 32, 2386-2395.

Noller AC, McEllistrem MC, Stine OC, Morris JG, Jr., Boxrud DJ, Dixon B and Harrison LH, 2003. Multilocus sequence typing reveals a lack of diversity among *Escherichia coli* O157:H7 isolates that are distinct by pulsed-field gel electrophoresis. J Clin Microbiol, 41, 675-679.

Octavia S and Lan R, 2006. Frequent recombination and low level of clonality within *Salmonella enterica* subspecies I. Microbiology, 152, 1099-1108.

On SL, 2013. Isolation, identification and subtyping of *Campylobacter*: where to from here? J Microbiol Methods, 95, 3-7.

Paddock Z, Shi X, Bai J and Nagaraja TG, 2012. Applicability of a multiplex PCR to detect O26, O45, O103, O111, O121, O145, and O157 serogroups of *Escherichia coli* in cattle feces. Vet Microbiol, 156, 381-388.

Paranthaman K, Haroon S, Latif S, Vinnyey N, de Souza V, Welfare W, Tahir M, Cooke E, Stone K, Lane C, Peters T and Puleston R, 2013. Emergence of a multidrug-resistant (ASSuTTm) strain of *Salmonella enterica* serovar Typhimurium DT120 in England in 2011 and the use of multiple-locus variable-number tandem-repeat analysis in supporting outbreak investigations. Foodborne Pathog Dis, 10, 850-855.

Pavlic M and Griffiths MW, 2009. Principles, applications, and limitations of automated ribotyping as a rapid method in food safety. Foodborne Pathog Dis, 6, 1047-1055.

Penny C, Mossong J, Walczak C, Devaux A, Collard D, Colin S, Fauvel B, Djabi F, Gadisseux L, Olinger C, Cauchie H-M and Ragimbeau C, 2013 O43b - Real-Time Surveillance of *Campylobacter* Linked to Detection in Environmental Waters and Wastewater. Abstracts of posters and oral presentations from the 17th International Workshop on *Campylobacter*, Helicobacter and Related Organisms, 15-19 September 2013, Aberdeen. Data suplemente of the Journal of Medical Microbiology. Available online at: http://jmm.sgmjournals.org/content/suppl/2013/08/30/62.Pt_9.DC1/CHROfinal.pdf.

Peters TM, Maguire C, Threlfall EJ, Fisher IS, Gill N and Gatto AJ, 2003. The Salm-gene project - a European collaboration for DNA fingerprinting for food-related salmonellosis. Euro Surveill, 8, 46-50.

Petersen RF, Litrup E, Larsson JT, Torpdahl M, Sorensen G, Muller L and Nielsen EM, 2011. Molecular characterization of *Salmonella* Typhimurium highly successful outbreak strains. Foodborne Pathog Dis, 8, 655-661.

Pezzoli L, Elson R, Little C, Fisher I, Yip H, Peters T, Hampton M, De Pinna E, Coia JE, Mather HA, Brown DJ, Nielsen EM, Ethelberg S, Heck M, de Jager C and Threlfall J, 2007. International outbreak of *Salmonella* Senftenberg in 2007. Euro Surveill, 12, E070614 070613.

Philippot L, Andersson SG, Battin TJ, Prosser JI, Schimel JP, Whitman WB and Hallin S, 2010. The ecological coherence of high bacterial taxonomic ranks. Nat Rev Microbiol, 8, 523-529.

Pike BL, Guerry P and Poly F, 2013. Global Distribution of Penner Serotypes: A Systematic Review. PLoS One, 8, e67375.

Pires SM, Evers EG, van Pelt W, Ayers T, Scallan E, Angulo FJ, Havelaar A and Hald T, 2009. Attributing the human disease burden of foodborne infections to specific sources. Foodborne Pathog Dis, 6, 417-424.

Pires SM and Hald T, 2010. Assessing the differences in public health impact of *Salmonella* subtypes using a bayesian microbial subtyping approach for source attribution. Foodborne Pathog Dis, 7, 143-151.

Pires SM, Kaesboher A, Spitznagel H, Whalstrom H, Nichols G, David J, Van Pelt W, Baumann A and T H, 2008. *Salmonella* source attribution in different European countries. Proceeding in FoodMicro 2008, Aberdeen, Scotland. Available online at: http://aberdeen.conference-services.net/programme.asp?conferenceID=1143&action=prog_categories (last accessed on 12/12/2013).

Pires SM, Knegt L and Hald T, 2011. Estimation of the relative contribution of different food and animal sources to human *Salmonella* infections in the European Union. Scientific/Technical Report submitted to EFSA. Question No EFSA-Q-2010-00685. Available online at: http://www.efsa.europa.eu/en/supporting/doc/184e.pdf (last accessed on 12/12/2013)

Poly F, Serichatalergs O, Schulman M, Ju J, Cates CN, Kanipes M, Mason C and Guerry P, 2011. Discrimination of major capsular types of *Campylobacter jejuni* by multiplex PCR. J Clin Microbiol, 49, 1750-1757.

Porwollik S and McClelland M, 2003. Lateral gene transfer in *Salmonella*. Microbes Infect, 5, 977-989.

Prendergast DM, O'Grady D, Fanning S, Cormican M, Delappe N, Egan J, Mannion C, Fanning J and Gutierrez M, 2011. Application of multiple locus variable number of tandem repeat analysis (MLVA), phage typing and antimicrobial susceptibility testing to subtype *Salmonella enterica* serovar Typhimurium isolated from pig farms, pork slaughterhouses and meat producing plants in Ireland. Food Microbiol, 28, 1087-1094.

Pritchard JK, Stephens M and Donnelly P, 2000. Inference of population structure using multilocus genotype data. Genetics, 155, 945-959.

Rabsch W, Hargis BM, Tsolis RM, Kingsley RA, Hinz KH, Tschape H and Baumler AJ, 2000. Competitive exclusion of *Salmonella* Enteritidis by *Salmonella* Gallinarum in poultry. Emerg Infect Dis, 6, 443-448.

Rabsch W, Mirold S, Hardt WD and Tschape H, 2002. The dual role of wild phages for horizontal gene transfer among *Salmonella* strains. Berl Munch Tierarztl Wochenschr, 115, 355-359.

Randall LP, Cooles SW, Osborn MK, Piddock LJ and Woodward MJ, 2004. Antibiotic resistance genes, integrons and multiple antibiotic resistance in thirty-five serotypes of *Salmonella enterica* isolated from humans and animals in the UK. J Antimicrob Chemother, 53, 208-216.

Ranieri ML, Shi C, Moreno Switt AI, den Bakker HC and Wiedmann M, 2013. Comparison of typing methods with a new procedure based on sequence characterization for *Salmonella* serovar prediction. J Clin Microbiol, 51, 1786-1797.

Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, Paxinos EE, Sebra R, Chin CS, Iliopoulos D, Klammer A, Peluso P, Lee L, Kislyuk AO, Bullard J, Kasarskis A, Wang S, Eid J, Rank D, Redman JC, Steyert SR, Frimodt-Moller J, Struve C, Petersen AM, Krogfelt KA, Nataro JP, Schadt EE and Waldor MK, 2011. Origins of the E. *coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. N Engl J Med, 365, 709-717.

Ribot EM, Fair MA, Gautom R, Cameron DN, Hunter SB, Swaminathan B and Barrett TJ, 2006. Standardization of pulsed-field gel electrophoresis protocols for the subtyping of *Escherichia coli* O157:H7, *Salmonella*, and *Shigella* for PulseNet. Foodborne Pathog Dis, 3, 59-67.

Ribot EM, Fitzgerald C, Kubota K, Swaminathan B and Barrett TJ, 2001. Rapid pulsed-field gel electrophoresis protocol for subtyping of *Campylobacter jejuni*. J Clin Microbiol, 39, 1889-1894.

Ripabelli G, McLauchin J and Threlfall EJ, 2000. Amplified fragment length polymorphism (AFLP) analysis of *Listeria monocytogenes*. Syst Appl Microbiol, 23, 132-136.

Rosef O, Kapperud G, Lauwers S and Gondrosen B, 1985. Serotyping of *Campylobacter jejuni*, *Campylobacter coli*, and *Campylobacter laridis* from domestic and wild animals. Appl Environ Microbiol, 49, 1507-1510.

Roukos DH, 2013. Integrated clinical genomics: new horizon for diagnostic and biomarker discoveries in cancer. Expert Rev Mol Diagn, 13, 1-4.

Sabat AJ, Budimir A, Nashev D, Sa-Leao R, van Dijl JM, Laurent F, Grundmann H, Friedrich AW and Esgem, 2013. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. Eurosurveillance, 18, 17-30.

Sahin O, Fitzgerald C, Stroika S, Zhao S, Sippy RJ, Kwan P, Plummer PJ, Han J, Yaeger MJ and Zhang Q, 2012. Molecular evidence for zoonotic transmission of an emergent, highly pathogenic *Campylobacter jejuni* clone in the United States. J Clin Microbiol, 50, 680-687.

Sandvang D, Aarestrup FM and Jensen LB, 2006. Characterisation of integrons and antibiotic resistance genes in Danish multiresistant *Salmonella enterica* Typhimurium DT104. FEMS microbiology letters, 160(1), 37-41.

Sangal V, Harbottle H, Mazzoni CJ, Helmuth R, Guerra B, Didelot X, Paglietti B, Rabsch W, Brisse S, Weill FX, Roumagnac P and Achtman M, 2010. Evolution and population structure of *Salmonella enterica* serovar Newport. J Bacteriol, 192, 6465-6476.

Sarwari AR, Magder LS, Levine P, McNamara AM, Knower S, Armstrong GL, Etzel R, Hollingsworth J and Morris JG, Jr., 2001. Serotype distribution of *Salmonella* isolates from food animals after slaughter differs from that of isolates found in humans. J Infect Dis, 183, 1295-1299.

Scheutz F, Teel LD, Beutin L, Pierard D, Buvens G, Karch H, Mellmann A, Caprioli A, Tozzoli R, Morabito S, Strockbine NA, Melton-Celsa AR, Sanchez M, Persson S and O'Brien AD, 2012. Multicenter evaluation of a sequence-based protocol for subtyping Shiga toxins and standardizing Stx nomenclature. J Clin Microbiol, 50, 2951-2963.

Segata N, Boernigen D, Tickle TL, Morgan XC, Garrett WS and Huttenhower C, 2013. Computational meta'omics for microbial community studies. Mol Syst Biol, 9, 666.

Shariat N, DiMarzio MJ, Yin S, Dettinger L, Sandt CH, Lute JR, Barrangou R and Dudley EG, 2013a. The combination of CRISPR-MVLST and PFGE provides increased discriminatory power for differentiating human clinical isolates of *Salmonella enterica* subsp. *enterica* serovar Enteritidis. Food Microbiol, 34, 164-173.

Shariat N, Kirchner MK, Sandt CH, Trees E, Barrangou R and Dudley EG, 2013b. Subtyping of *Salmonella enterica* serovar Newport outbreak isolates by CRISPR-MVLST and determination of the relationship between CRISPR-MVLST and PFGE results. J Clin Microbiol, 51, 2328-2336.

Sharma VK and Bearson SM, 2013. Evaluation of the impact of quorum sensing transcriptional regulator SdiA on long-term persistence and fecal shedding of *Escherichia coli* O157:H7 in weaned calves. Microb Pathog, 57, 21-26.

Shen J, Rump L, Zhang Y, Chen Y, Wang X and Meng J, 2013. Molecular subtyping and virulence gene analysis of *Listeria monocytogenes* isolates from food. Food microbiology, 35(1), 58-64.

Sheppard SK, McCarthy ND, Falush D and Maiden MC, 2008. Convergence of *Campylobacter* species: implications for bacterial evolution. Science, 320, 237-239.

Sheppard SK, Dallas JF, Strachan NJ, MacRae M, McCarthy ND, Wilson DJ, Gormley FJ, Falush D, Ogden ID, Maiden MC and Forbes KJ, 2009. *Campylobacter* genotyping to determine the source of human infection. Clin Infect Dis, 48, 1072-1078.

Sheppard SK, Colles FM, McCarthy ND, Strachan NJ, Ogden ID, Forbes KJ, Dallas JF and Maiden MC, 2011a. Niche segregation and genetic structure of *Campylobacter jejuni* populations from wild and agricultural host species. Mol Ecol, 20, 3484-3490.

Sheppard SK, McCarthy ND, Jolley KA and Maiden MC, 2011b. Introgression in the genus *Campylobacter*: generation and spread of mosaic alleles. Microbiology, 157, 1066-1074.

Sheppard SK, Didelot X, Jolley KA, Darling AE, Pascoe B, Meric G, Kelly DJ, Cody A, Colles FM, Strachan NJ, Ogden ID, Forbes K, French NP, Carter P, Miller WG, McCarthy ND, Owen R, Litrup E, Egholm M, Affourtit JP, Bentley SD, Parkhill J, Maiden MC and Falush D, 2013. Progressive genome-wide introgression in agricultural *Campylobacter coli*. Mol Ecol, 22, 1051-1064.

Sheppard SK, Didelot X, Meric G, Torralbo A, Jolley KA, Kelly DJ, Bentley SD, Maiden MC, Parkhill J and Falush D, 2013b. Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in *Campylobacter*. Proc Natl Acad Sci U S A, 110, 11923-11927.

Siebor E and Neuwirth C, 2013. Emergence of *Salmonella* genomic island 1 (SGI1) among Proteus mirabilis clinical isolates in Dijon, France. J Antimicrob Chemother, 68, 1750-1756.

Slauch JM, Mahan MJ, Michetti P, Neutra MR and Mekalanos JJ, 1995. Acetylation (O-factor 5) affects the structural and immunological properties of Salmonella Typhimurium lipopolysaccharide O antigen. Infect Immun, 63, 437-441.

Smid JH, Mughini Gras L, de Boer AG, French NP, Havelaar AH, Wagenaar JA and van Pelt W, 2013. Practicalities of using non-local or non-recent multilocus sequence typing data for source attribution in space and time of human campylobacteriosis. PLoS One, 8, e55029.

Soborg B, Lassen SG, Muller L, Jensen T, Ethelberg S, Molbak K and Scheutz F, 2013. A verocytotoxin-producing *E. coli* outbreak with a surprisingly high risk of haemolytic uraemic syndrome, Denmark, September-October 2012. Euro Surveill, 18.

Son I, Zheng J, Keys CE, Zhao S, Meng J and Brown EW, 2013. Analysis of pulsed field gel electrophoresis profiles using multiple enzymes for predicting potential source reservoirs for strains of *Salmonella* Enteritidis and *Salmonella* Typhimurium isolated from humans. Infect Genet Evol, 16, 226-233.

Spratt BG, Hanage WP, Li B, Aanensen DM and Feil EJ, 2004. Displaying the relatedness among isolates of bacterial species -- the eBURST approach. FEMS Microbiol Lett, 241, 129-134.

SSI and DTU (Statens Serum Institut and Technical University of Denmark (National Veterinary Institute and National Food Institute).), 2013. DANMAP 2012 - Use of antimicrobial agents and occurrence of antimicrobial resistance in bacteria from food animals, food and humans in Denmark. Available online at: http://www.danmap.org/Downloads/~/media/Projekt%20sites/Danmap/DANMAP%20reports/DANMAP%202012/Danmap_2012.ashx (last acccessed on 12/12/2013).

Strachan NJ, Gormley FJ, Rotariu O, Ogden ID, Miller G, Dunn GM, Sheppard SK, Dallas JF, Reid TM, Howie H, Maiden MC and Forbes KJ, 2009. Attribution of *Campylobacter* infections in northeast Scotland to specific sources by use of multilocus sequence typing. J Infect Dis, 199, 1205-1208.

Struelens MJ, 1998. Molecular epidemiologic typing systems of bacterial pathogens: current issues and perspectives. Mem Inst Oswaldo Cruz 93, 581-585.

Struelens MJ and Brisse S, 2013. From molecular to genomic epidemiology: transforming surveillance and control of infectious diseases. Euro Surveill, 18, 20386.

Suez J, Porwollik S, Dagan A, Marzel A, Schorr YI, Desai PT, Agmon V, McClelland M, Rahav G and Gal-Mor O, 2013. Virulence gene profiling and pathogenicity characterization of non-typhoidal *Salmonella* accounted for invasive disease in humans. PLoS One, 8, e58449.

Swaminathan B, Barrett TJ, Hunter SB and Tauxe RV, 2001. PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. Emerg Infect Dis, 7, 382-389.

Swaminathan B, Gerner-Smidt P, Ng LK, Lukinmaa S, Kam KM, Rolando S, Gutierrez EP and Binsztein N, 2006. Building PulseNet International: an interconnected system of laboratory networks to facilitate timely public health recognition and response to foodborne disease outbreaks and emerging foodborne diseases. Foodborne Pathog Dis, 3, 36-50.

Szymanski CM and Gaynor EC, 2012. How a sugary bug gets through the day: recent developments in understanding fundamental processes impacting *Campylobacter jejuni* pathogenesis. Gut Microbes, 3, 135-144.

Tang L, Li Y, Deng X, Johnston RN, Liu GR and Liu SL, 2013. Defining natural species of bacteria: clear-cut genomic boundaries revealed by a turning point in nucleotide sequence divergence. BMC Genomics, 14, 489.

Tankouo-Sandjong B, Kinde H and Wallace I, 2012. Development of a sequence typing scheme for differentiation of *Salmonella* Enteritidis strains. FEMS Microbiol Lett, 331, 165-175.

Taylor EV, Herman KM, Ailes EC, Fitzgerald C, Yoder JS, Mahon BE and Tauxe RV, 2012. Common source outbreaks of *Campylobacter* infection in the USA, 1997-2008. Epidemiol Infect, 141, 987-996.

Toyofuku H, Pires SM and Hald T, 2011. *Salmonella* source attribution in Japan by a microbial subtyping approach. EcoHealth (ISSN: 1612-9202), 7 (Suppl.1): S22-S23.

Turabelidze G, Lawrence SJ, Gao H, Sodergren E, Weinstock GM, Abubucker S, Wylie T, Mitreva M, Shaikh N, Gautom R and Tarr PI, 2013. Precise Dissection of an *Escherichia coli* O157:H7 Outbreak by Single Nucleotide Polymorphism Analysis. J Clin Microbiol, 51, 3950-3954.

Underwood AP, Dallman T, Thomson NR, Williams M, Harker K, Perry N, Adak B, Willshaw G, Cheasty T, Green J, Dougan G, Parkhill J and Wain J, 2013. Public health value of next-generation DNA sequencing of enterohemorrhagic *Escherichia coli* isolates from an outbreak. J Clin Microbiol, 51, 232-237.

Valentin L, Sharp H, Appel B and Kasbohrer A, 2013. Source attribution of foodborne ESBL-*E. coli* in Germany. Abstract Ref. RR11. Med-Vet-Net Association International Scientific Conference. DTU, Lyngby, Denmark, 24-25 June 2013. Available online at: http://www.medvetnet2013.eu/fileadmin/filer/David/MedVetNet2013/MVN2013ConfAbstractBook_v1.0.pdf (last accessed on 16/12/2013).

Valkenburgh S, van Oosterom R, Stenvers O, Aalten M, Braks M, Schimmer B, Van De Giessen AW, Van Pelt W and Langelaar M, 2007. Zoonoses and zoonotic agents in humans, food, animals and feed in The Netherlands 2003-2006. Available online at: http://www.rivm.nl/bibliotheek/rapporten/330152001.pdf (last accessed on 13/12/2013).

van Alphen LB, Bleumink-Pluym NM, Rochat KD, van Balkom BW, Wosten MM and van Putten JP, 2008. Active migration into the subcellular space precedes *Campylobacter jejuni* invasion of epithelial cells. Cell Microbiol, 10, 53-66.

van Belkum A, Struelens M, de Visser A, Verbrugh H and Tibayrenc M, 2001. Role of genomic typing in taxonomy, evolutionary genetics, and microbial epidemiology. Clin Microbiol Rev, 14, 547-560.

van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, Fry NK, Fussing V, Green J, Feil E, Gerner-Smidt P, Brisse S and Struelens M, 2007. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clin Microbiol Infect, 13 Suppl 3, 1-46.

Van Pelt W, Van De Giessen AW, Leeuwen WJ, Wannet W, Henken AM and Evers EG, 1999. Oorsprong, omvang en kosten van humane salmonellose. Deel 1. Oorsprong van humane salmonellose met betrekking tot varken, rund, kip, ei en overige bronnen. Infectieziekten Bulletin, 240-243.

Vitullo M, Grant KA, Sammarco ML, Tamburro M, Ripabelli G and Amar CF, 2013. Real-time PCRs assay for serogrouping *Listeria monocytogenes* and differentiation from other *Listeria* spp. Mol Cell Probes, 27, 68-70.

Waddell SJ, von Kamp A, Klamt S and Neyrolles O 2013. Host–Pathogen Interactions. In Systems Biology of Tuberculosis (pp. 107-126). Springer New York.

Wahlstrom H, Andersson Y, Plym-Forshell L and Pires SM, 2010. Source attribution of human *Salmonella* cases in Sweden. Epidemiol Infect, 139, 1246-1253.

Wang L, Wakushima M, Aota T, Yoshida Y, Kita T, Maehara T, Ogasawara J, Choi C, Kamata Y, Hara-Kudo Y and Nishikawa Y, 2013. Specific properties of enteropathogenic *Escherichia coli* isolates from diarrheal patients and comparison to strains from foods and fecal specimens from cattle, swine, and healthy carriers in Osaka City, Japan. Appl Environ Microbiol, 79, 1232-1240.

Wassenaar TM, Fernandez-Astorga A, Alonso R, Marteinsson VT, Magnusson SH, Kristoffersen AB and Hofshagen M, 2009. Comparison of *Campylobacter* fla-SVR genotypes isolated from humans and poultry in three European regions. Lett Appl Microbiol, 49, 388-395.

Wiedmann M, Bruce JL, Knorr R, Bodis M, Cole EM, McDowell CI, McDonough PL and Batt CA, 1996. Ribotype diversity of *Listeria monocytogenes* strains associated with outbreaks of listeriosis in ruminants. J Clin Microbiol, 34, 1086-1090.

Wilson DJ, Gabriel E, Leatherbarrow AJ, Cheesbrough J, Gee S, Bolton E, Fox A, Fearnhead P, Hart CA and Diggle PJ, 2008. Tracing the source of campylobacteriosis. PLoS Genet, 4, e1000203.

Wilson DJ, Gabriel E, Leatherbarrow AJ, Cheesbrough J, Gee S, Bolton E, Fox A, Hart CA, Diggle PJ and Fearnhead P, 2009. Rapid evolution and the importance of recombination to the gastroenteric pathogen *Campylobacter jejuni*. Mol Biol Evol, 26, 385-397.

Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM and Larsen MV, 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother, 67, 2640-2644.

Zankari E, Hasman H, Kaas RS, Seyfarth AM, Agerso Y, Lund O, Larsen MV and Aarestrup FM, 2013. Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. J Antimicrob Chemother, 68, 771-777.

Zhang J, Cao G, Xu X, Jin H, Zhang Q, Chen J, Yang X, Pan H, Zhang X, Allard M, Brown E and Meng J, 2013. Whole-Genome Sequences of Four *Salmonella enterica* erotype Newport Strains from Humans. Genome Announc, 1 (3).

APPENDICES

Appendix A. Summary of the evaluation of molecular typing methods for *Salmonella* spp., STEC, *Listeria monocytogenes* and *Campylobacter* spp.

Table 1 below summarises the results of the evaluation of the molecular typing methods for the different pathogens considered in this Opinion as per the criteria described in detail in chapter three above (i.e. discriminatory capability, reproducibility and aspects of international harmonisation).

Table 1: Summary of the results of the evaluation of the molecular typing methods considered in this Opinion for the different pathogens. The detailed evaluation can be found in chapter 3 in the main body of this document.

Method	Primary/secondary method?	Pathogen	How much and how accurately samples the genome?	Discriminatory capability	Reproducibility (between labs)	Current international harmonisation	Capability for future international harmonisation
Molecular serotyping	Primary	*Salmonella* spp.	Depends on chosen methodology	Low to moderate, usually too low for outbreak investigations	Depends on methodology	• Standard Operational Procedures (SOP) - no • External Quality Assurance (EQA) - no • Nomenclature (NOM) - yes, follows same nomenclature as traditional serotyping. • Data management tools (DMT) - yes	• SOP - yes • EQA - yes • NOM - yes • DMT - NA[(a)]
	Primary	*Campylobacter* spp.	Samples limited regions of the genome	Low to moderate, usually too low for outbreak investigations	Moderate	• SOP - no • EQA - no • NOM - yes • DMT - yes	• SOP - yes • EQA - yes • NOM – NA • DMT - NA
	Primary	STEC	Depends on chosen methodology	Low to moderate, usually too low for outbreak investigations	Depends on methodology	• SOP - no • EQA - no • NOM - yes • DMT - yes[(b)]	• SOP - yes • EQA - yes • NOM - NA • DMT - NA
	Primary	*Listeria monocytogenes*	Samples limited regions of the genome	Low to moderate, usually too low for outbreak investigations	Moderate to high	• SOP - no • EQA - no • NOM - yes • DMT - yes	• SOP - yes • EQA - yes • NOM - NA • DMT - NA

Method	Primary/secondary method?	Pathogen	How much and how accurately samples the genome?	Discriminatory capability	Reproducibility (between labs)	Current international harmonisation	Capability for future international harmonisation
Restriction fragment length polymorphism (RLFP)	Primary	For all four pathogens	Samples a limited region of the genome	Low to high (depends on technology used)	Low to high (depends on platform)	• SOP - yes, automated • EQA - yes, automated • NOM - yes, automated • DMT - yes, automated	NA
Pulsed-field gel electrophoresis (PFGE)	Primary/secondary	*Salmonella* spp.	Samples whole genome, but achieves this by cutting DNA at limited restriction sites (usually 6bp regions)	Variable, depends on serovar; useful for outbreak investigation: discriminates outbreak/non - outbreak strains.	Moderate to high	• SOP - yes • EQA - yes • NOM - yes • DMT - yes	NA
	Secondary	*Campylobacter* spp.	Samples whole genome, but achieves this by cutting DNA at limited restriction sites (usually 6bp regions).	High, useful for outbreak investigation: discriminates outbreak/non - outbreak strains.	Low to moderate	• SOP - yes • EQA - yes • NOM - yes • DMT - yes	NA
	Secondary	STEC	Samples whole genome, but achieves this by cutting DNA at limited restriction sites (usually 6bp regions)	High, useful for outbreak investigation: discriminates outbreak/non - outbreak strains.	Moderate to high	• SOP - yes • EQA - yes • NOM - yes • DMT - yes	NA
	Secondary	*Listeria moncytogenes*	Samples whole genome, but achieves this by cutting DNA at limited restriction sites (usually 6bp regions)	High, useful for outbreak investigation: discriminates outbreak/non - outbreak strains.	Moderate to high	• SOP - yes • EQA - yes • NOM - yes • DMT - yes	NA

Method	Primary/secondary method?	Pathogen	How much and how accurately samples the genome?	Discriminatory capability	Reproducibility (between labs)	Current international harmonisation	Capability for future international harmonisation
Specific gene characterisation	Primary	For all four pathogens	Samples selected regions of the genome (usually an array of functional genes)	Variable, it depends on selected markers and intended use.	Moderate	• SOP - no • EQA - no • NOM - no • DMT – yes[(b)]	• SOP - yes • EQA - yes • NOM - yes • DMT - yes
Multiple loci variable tandem repeat analysis (MLVA)	Primary/secondary	*Salmonella* spp.	Samples limited regions of the genome (VNTR loci)	Variable, it depends on serovar. High for *S.* Typhimurium.	Moderate to high. High for *S.* Typhimurium.	For *S.* Typhimurium: • SOP - yes • EQA - yes • NOM - yes • DMT - yes	For other *Salmonella* serovars: • SOP - yes • EQA - yes • NOM - yes • DMT - yes
	Primary/secondary	STEC	Samples limited regions of the genome (VNTR loci)	Variable, it depends on serotype. High for O157:H7	Moderate to high.	• SOP - no • EQA - no • NOM - no • DMT – yes[(b)]	• SOP - yes • EQA -yes • NOM - yes • DMT - yes
	Primary/secondary	*Listeria monocytogenes*	Samples limited regions of the genome (VNTR loci)	Variable, it depends on serotype. High for serotypes associated with human infections of *L. monocytogenes*	Moderate to high	• SOP - no • EQA - no • NOM - no • DMT – yes[(b)]	• SOP - yes • EQA - yes • NOM - yes • DMT - yes
Single locus sequence typing (SLST)	Secondary	For all four pathogens	Samples limited regions of the genome (usually single genes)	Variable, depends on assay and pathogen	High	Depends on assay, since this method can involve any gene. It is not feasible to make a general statement.	NA

Method	Primary/secondary method?	Pathogen	How much and how accurately samples the genome?	Discriminatory capability	Reproducibility (between labs)	Current international harmonisation	Capability for future international harmonisation
Multi locus sequence typing (MLST)	Primary/secondary	For all four pathogens	Samples limited regions of the genome (selected housekeeping genes)	Moderate (Usually not high enough for outbreak investigations involving major food-borne pathogens)	High	• SOP - yes • EQA - yes • NOM - yes • DMT - yes	NA
Whole genome mapping (optical mapping)	Primary/secondary	For all four pathogens	Samples restriction site polymorphisms throughout the genome	Unknown, probably high	High for automated system	• SOP - no • EQA - no • NOMT - no • DMT - yes	• SOP - yes • EQA - yes • NOM - yes • DMT – NA
Whole genome sequencing (WGS)	Primary	For all four pathogens	Samples the whole genome as well as extrachromosomal DNA. Different technologies do display various limitations e.g. in sequencing repeated DNA motifs.	High	High	• SOP – no • EQA - no • NOM - yes • DMT - yes	• SOP - yes • EQA - yes • NOM - yes • DMT - NA

(a) NA=not applicable
(b) most of the available DNA-software management tools can be used

GLOSSARY

Accessory genome: those genes shared among, but not present in all, members of a defined subset of bacteria.

Bottlenecking: the reduction in diversity of a microbial population caused by very few members of a population leaving descendants. This term is normally used to imply a stochastic process, i.e. the survival of cells from one generation to the next being due to chance.

Clade: monophyletic, meaning it contains one ancestor (which can be an organism, a population, or a species) and all its descendants.

Clonal population: a population where all the members are related to a single ancestor, with evolution proceeding solely by mutation and vertical genetic transfer, with accompanying diversity reduction events.

Congruence: the same phylogenetic signal being recorded at two or more loci.

Core genome: those parts of the genome shared by all members of a defined subset of bacteria.

Diversity reduction: the loss of genetic variation from the population.

Epigenetics: study of the heritable changes in the activity (expression) of genes that are not due to changes in the DNA sequence.

Genetic drift: the genes that are passed from one generation of organism to the next one due to chance and randomness. This is one of the mechanisms involved in bacterial evolution together with others like natural selection and mutation.

Homology: any similarity between characteristics that is due to their shared ancestry.

Homoplasy: a trait (genetic or morphological) that is shared by two or more group of populations of organisms due to convergence, parallelism or reversal and not common descent.

Horizontal gene transfer: the passing of genetic information among cells that do not necessarily share a common parent by processes other than descent.

Hypermutation: spontaneous mutations at high rates that may occur in microorganisms.

Lineage: a group of bacteria all of which share an ancestor, usually used to define clonal subgroups within bacterial populations.

Linkage disequilibrium: the non-random association of allelic variants in populations of organisms, typical of asexual i.e. non-recombining, populations.

Meroclone: group of microorganisms originating from single cell that have started to diversify by recombination.

Metagenomics: study of the genetic material directly recovered from environmental samples so called 'metagenomes'. Also known as environmental genomics or ecogenomics.

Mosaicism: presence of two genetic information with different evolutionary histories in the same isolate or gene.

Non-clonal population: a population in which the spread of genetic variation is dominated by horizontal gene transfer.

Pan genome: the total complement of genes (core plus accessory genome) that is available to a defined subset of bacteria.

Periodic selection: the selection of particular variants. Like bottlenecking, this results in diversity reduction, but in this case it is cause by selection i.e. the survival of particular variants to the next generation because they are more fit than others in the population.

Pipeline: computational algorithms for detecting and interpreting variants from alignment of genomic sequences.

Strain: genetic of phenotypic subtype of a microorganism often defined for epidemiological purposes.

Subspecies: in bacterial taxonomy subgroups of a specie that differ in their phenotypic or genotypic characteristics.

Subtype: discrimination of bacteria below the level of species.

Synteny: the organisation of loci in the same chromosome, which is maintained either in a species or among isolates.

Tree-like phylogeny: a bifurcating representation of the genealogy of a group of organisms.

Vertical genetic transfer: the passing on of genetic material by descent from parent cell to progeny.

ABBREVIATIONS

CGH Comparative Genomic Hybridization

CRISPR Clustered Regularly Interspaced Short Palindromic Repeats

DNA Deoxyribonucleic acid

EQA External Quality Assurance system

MLST Multi-locus Strain Typing

MLVST Multi-Virulence-Locus Sequence Typing

MLVA Multi-Locus Variable number tandem repeat Analysis

NGS Next Generation Sequencing

OBGS Octamer-Based Genome Scanning

PCR Polymerase Chain Reaction

PFGE Pulsed Field Gel Electrophoresis

RLFP Restriction Fragment Length Polymorphism

RNA Ribonucleic acid

rRNA Ribosomal ribonucleic acid

SLST Single Locus Sequence Typing

SOP Standard Operational Procedure

STEC Shiga-like toxin-producing *Escherichia coli*

SVR Short Variable Region

VNTR Variable Number Tandem Repeat

VTEC Verotoxin-producing *E. coli*

WGM Whole Genome Mapping

WGS Whole Genome Sequencing